

THIAGO VIEIRA DA SILVA

COMO VARIÁVEIS ECONÔMICAS INFLUENCIAM O COMPORTAMENTO
DO IBOVESPA SOB UMA PERSPECTIVA FUTURA DO MERCADO

MONOGRAFIA APRESENTADA À ESCOLA
POLITÉCNICA DA UNIVERSIDADE DE SÃO
PAULO, PARA OBTENÇÃO DO TÍTULO DE
MBA EM ENGENHARIA FINANCEIRA

ÁREA DE CONCENTRAÇÃO:
ENGENHARIA FINANCEIRA

ORIENTADOR: PROF. DR CLAUDIO GARCIA

SÃO PAULO

2015

RESUMO

O objetivo desse trabalho é elaborar um estudo sobre a relação do IBOVESPA com o movimento histórico de diversas variáveis, que podem influenciar o mercado acionário brasileiro.

A principal hipótese é que existe uma relação entre diversas variáveis de âmbito econômico e o índice IBOVESPA. Portanto, pressupõe-se que este assunto é de interesse de instituições financeiras, tais como corretoras e bancos.

Uma vez determinado como o IBOVESPA se comporta, também seria possível estender esta estimativa aos ativos que compõem uma carteira e então garantir um melhor retorno, considerando uma perspectiva futura do mercado.

Palavras-chave: IBOVESPA, Regressão Múltipla, Previsão

ABSTRACT

The goal of this work is to prepare a study on the Ibovespa's relation to the historical movement of several variables that can influence the Brazilian stock Market

The main hypothesis is that there is a relationship between several variables of economic scope and Ibovespa index. Therefore, it is assumed that this subject is of interest to financial institutions, such as brokerages and banks.

Once determined how Ibovespa behaves, it would also be possible to extend this estimate to assets that are part of a portfolio and then ensure a better return, considering a future market perspective.

Keywords: IBOVESPA, Multiple Regression, Forecast

LISTA DE ILUSTRAÇÕES

Figura 1 – Série histórica das cotações do IBOVESPA em 2014 (UOL).....	28
Figura 2 – Série histórica dos retornos do índice IBOVESPA em 2014 (UOL).....	29
Figura 3 – Um processo estocástico interpretado como uma família de variáveis aleatórias (MORETTIN e TOLOI, 2006, p. 20).....	30
Figura 4 – Gráfico de dispersão IBOVESPA x Dólar americano.....	32
Figura 5 – Gráficos das séries temporais IBOVESPA, CAC40 e dispersão IBOVESPA x CAC40.....	41
Figura 6 – Gráficos das séries temporais IBOVESPA, DAX30 e dispersão IBOVESPA x DAX30.....	42
Figura 7 – Gráficos das séries temporais IBOVESPA, DJI e dispersão IBOVESPA x DJI.....	42
Figura 8 – Gráficos das séries temporais IBOVESPA, FTSEMIB e dispersão IBOVESPA x FTSEMIB.....	42
Figura 9 – Gráficos das séries temporais IBOVESPA, HSENG e dispersão IBOVESPA x HSENG.....	42
Figura 10 – Gráficos das séries temporais IBOVESPA, IBEX35 e dispersão IBOVESPA x IBEX35.....	43
Figura 11 – Gráficos das séries temporais IBOVESPA, IPCMEX e dispersão IBOVESPA x IPCMEX.....	43
Figura 12 – Gráficos das séries temporais IBOVESPA, Merval e dispersão IBOVESPA x Merval.....	43
Figura 13 – Gráficos das séries temporais IBOVESPA, NASDAQC e dispersão IBOVESPA x NASDAQC.....	43
Figura 14 – Gráficos das séries temporais IBOVESPA, NIKKEI225 e dispersão IBOVESPA x NIKKEI225.....	44
Figura 15 – Gráficos das séries temporais IBOVESPA, SHNGYC e dispersão IBOVESPA x SHNGYC.....	44
Figura 16 – Gráficos das séries temporais IBOVESPA, SMI e dispersão IBOVESPA x SMI.....	44

Figura 17 – Gráficos das séries temporais IBOVESPA, SP500 e dispersão IBOVESPA x SP500.....	44
Figura 18 – Gráficos das séries temporais IBOVESPA, SELIC e dispersão IBOVESPA x SELIC.....	45
Figura 19 – Gráficos das séries temporais IBOVESPA, DÓLAR e dispersão IBOVESPA x DÓLAR.....	45
Figura 20 – Gráficos das séries temporais IBOVESPA, EURO e dispersão IBOVESPA x EURO.....	45
Figura 21 – Gráficos das séries temporais IBOVESPA, LIBRA e dispersão IBOVESPA x LIBRA.....	45
Figura 22 – Gráficos das séries temporais IBOVESPA, PESO ARGENTINO e dispersão IBOVESPA x PESO ARGENTINO.....	46
Figura 23 – Gráficos das séries temporais IBOVESPA, IENE e dispersão IBOVESPA x IENE.....	46
Figura 24 – Gráficos das séries temporais IBOVESPA, OURO e dispersão IBOVESPA x OURO.....	46
Figura 25 – Gráficos das séries temporais IBOVESPA, SOJA e dispersão IBOVESPA x SOJA.....	46
Figura 26 – Gráficos das séries temporais IBOVESPA, CAFÉ e dispersão IBOVESPA x CAFÉ.....	47
Figura 27 – Gráficos das séries temporais IBOVESPA, PETRÓLEO e dispersão IBOVESPA x PETRÓLEO.....	47
Figura 28 – Gráficos das séries temporais IBOVESPA, TRIGO e dispersão IBOVESPA x TRIGO.....	47
Figura 29 – Gráficos das séries temporais IBOVESPA, AÇÚCAR e dispersão IBOVESPA x AÇÚCAR.....	47
Figura 30 – Gráficos das séries temporais IBOVESPA, EMBI+Risco Brasil e dispersão IBOVESPA x EMBI+ Risco Brasil.....	48
Figura 31 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (18).	54
Figura 32 – Resíduos do modelo (18) em relação ao tempo.....	56

Figura 33 – Histograma dos resíduos do modelo (18).....	56
Figura 34 – Resíduos versus valores ajustados do modelo (18).....	57
Figura 35 – Gráfico normal de probabilidade, ou <i>quantil quantil</i> , dos resíduos do modelo (18).....	57
Figura 36 – Gráficos de dispersão para comparar a qualidade das respostas e os valores observados entre os períodos de ajuste e teste do modelo (18).....	58
Figura 37 – Gráfico da função de autocorrelação dos resíduos do modelo (18).....	58
Figura 38 – Gráfico da função de autocorrelação parcial dos resíduos do modelo (18).....	59
Figura 39 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (18) com a adição do modelo auto regressivo dos resíduos (4-2).....	60
Figura 40 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (18) antes e depois da adição do modelo auto regressivo dos resíduos (4-2).....	60
Figura 41 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (17).....	63
Figura 42 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (17).....	64
Figura 43 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.....	64
Figura 44 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (17).....	65
Figura 45 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (17) com a adição do modelo auto regressivo dos resíduos (4-4).....	66
Figura 46 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (17) antes e depois da adição do modelo auto regressivo dos resíduos (4-4).....	66
Figura 47 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (12).....	69
Figura 48 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (12).....	69

Figura 49 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.....	70
Figura 50 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (12).	70
Figura 51 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (12) com a adição do modelo auto regressivo dos resíduos (4-6).....	71
Figura 52 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (12) antes e depois da adição do modelo auto regressivo dos resíduos (4-6).	71
Figura 53 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (9).	73
Figura 54 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (9).	74
Figura 55 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos do modelo (9).....	74
Figura 56 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (9).	75
Figura 57 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (9) com a adição do modelo auto regressivo dos resíduos (4-8).....	76
Figura 58 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (9) antes e depois da adição do modelo auto regressivo dos resíduos (4-8).	76
Figura 59 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (7).	79
Figura 60 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (7).	80
Figura 61 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos do modelo (7).....	80
Figura 62 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (7).	81
Figura 63 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-11)...82	

Figura 64 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-11).....	82
Figura 65 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-12)...	83
Figura 66 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-12).....	83
Figura 67 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-13)...	84
Figura 68 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-13).....	84
Figura 69 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-14)...	85
Figura 70 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-14).....	85
Figura 71 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (7).....	88
Figura 72 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (9-2).....	89
Figura 73 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.....	89
Figura 74 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (9-2)	89
Figura 75 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (9-2) mais modelo dos resíduos.	90
Figura 76 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (9-2) antes e depois da soma dos resíduos.....	91

LISTA DE TABELAS

Tabela 1 – Índices de ações selecionados para o estudo do modelo de previsão do IBOVESPA.....	20
Tabela 2 – Moedas selecionadas para o estudo do modelo de previsão do IBOVESPA.....	24
Tabela 3 – Commodities selecionadas para o estudo do modelo de previsão do IBOVESPA.....	25
Tabela 4 – Correlações dos vinte e seis índices com o IBOVESPA, calculados através do Excel.....	49
Tabela 5 – Resultado da função <i>Best Subsets</i> do Minitab para o conjunto de vinte e seis variáveis.....	50
Tabela 6 – Comparação dos coeficientes <i>R-Sq</i> e <i>R-Sq (adj)</i> da função <i>Best Subsets</i> do Minitab.....	51
Tabela 7 – Modelo de regressão múltipla com 18 variáveis gerado pelo Minitab.....	52
Tabela 8 – Referência de valores para interpretação do VIF (FROST, 2014).....	53
Tabela 9 – Valores iniciais da série histórica do IBOVESPA, os valores ajustados pelo modelo (18) e os resíduos calculados.....	55
Tabela 10 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (18).....	59
Tabela 11 - Correlações dos vinte índices mais relevantes com o IBOVESPA.....	61
Tabela 12 – Resultado da função <i>Best Subsets</i> do Minitab para o conjunto de vinte variáveis.....	61
Tabela 13 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo (17) variáveis.....	62
Tabela 14 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (17).....	65
Tabela 15 – Correlações dos quinze índices mais relevantes com o IBOVESPA.....	67
Tabela 16 – Resultado da função <i>Best Subsets</i> do Minitab para o conjunto das quinze variáveis.....	67
Tabela 17 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo com 12 variáveis.....	68
Tabela 18 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (12).....	70

Tabela 19 – Correlações dos dez índices mais relevantes com o IBOVESPA.	72
Tabela 20 – Resultado da função <i>Best Subsets</i> do Minitab 17 para o conjunto das dez variáveis.	72
Tabela 21 – Coeficientes da regressão múltipla calculados pelo Minitab 17 para o modelo com 9 variáveis.	73
Tabela 22 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (12)..	75
Tabela 23 – Comparação do valor do parâmetro VIF (multicolinearidade) das variáveis do modelo (9), com os modelos anteriores que também contém as mesmas variáveis.	77
Tabela 24 – Resultado da função <i>Best Subsets</i> do Minitab para o conjunto de dez variáveis, porém com o IBOVESPA deslocado cinco dias.	78
Tabela 25 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo com 7 variáveis.	78
Tabela 26 – Comparação do parâmetro VIF entre os modelos (7) e (9).	79
Tabela 27 – Modelo auto regressivo AR(5) para a série dos resíduos do modelo (12)..	81
Tabela 28 – Comparação das correlações de cada variável após o atraso em relação ao IBOVESPA.	86
Tabela 29 – Lista das dez variáveis mais relevantes em termos de correlação com o IBOVESPA.	87
Tabela 30 – Resultado da função <i>Best Subsets</i> para o conjunto das dez variáveis da Tabela 24.	87
Tabela 31 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo (9-2).....	88
Tabela 32 – Modelo auto regressivo AR(2) para a série dos resíduos do modelo (9-2).	90
Tabela 33 - Comparação de R ² entre os modelos.....	92

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	DEFINIÇÃO DO PROBLEMA	14
1.2	OBJETIVO DO TRABALHO.....	15
1.3	JUSTIFICATIVA DO TRABALHO	16
1.4	ESTRUTURA DO TRABALHO.....	16
1.5	LIMITAÇÕES DO TRABALHO.....	17
2	ESCOLHAS DAS VARIÁVEIS DE ENTRADA DO MODELO	18
2.1	PERÍODO E FONTES PARA COLETA DOS DADOS	19
2.2	TRATAMENTO E PREPARAÇÃO DOS DADOS.....	19
2.3	ÍNDICES DE AÇÕES	19
2.3.1	CAC40	20
2.3.2	DAX30	21
2.3.3	DJI	21
2.3.4	FTSEMIB	21
2.3.5	HSENG	21
2.3.6	IBEX35	22
2.3.7	IPCMEX	22
2.3.8	MERVAL.....	22
2.3.9	NASDAQC	22
2.3.10	NIKKEI225.....	23
2.3.11	SHNGYC	23
2.3.12	SMI	23
2.3.13	SP500.....	23
2.4	MOEDAS.....	24

2.5	MEDIDA RISCO BRASIL	24
2.6	COMMODITIES	25
2.7	TAXA DE JUROS FUTUROS	26
3	REFERENCIAL TEÓRICO	27
3.1	TIPOS DE DADOS.....	27
3.2	RETORNOS.....	28
3.3	PROCESSOS ESTOCÁSTICOS	30
3.4	FUNÇÃO DE DENSIDADE DE PROBABILIDADE E DISTRIBUIÇÃO DE PROBABILIDADE	31
3.5	REGRESSÃO LINEAR.....	31
3.6	MODELO DE REGRESSÃO COM DUAS VARIÁVEIS.....	32
3.7	COEFICIENTE DE DETERMINAÇÃO R^2	35
3.8	REGRESSÃO MÚLTIPLA.....	35
3.9	MODELOS AUTO REGRESSIVOS	37
3.10	MODELOS DE MÉDIAS MÓVEIS	37
3.11	MODELOS ARMA.....	38
3.12	IDENTIFICAÇÃO DE MODELOS ARMA	38
3.12.1	Autocorrelação	38
3.12.2	Autocorrelação Parcial	39
3.12.3	Escolha do Modelo AR(1), MA(1) e ARMA(1,1).....	39
3.13	MULTICOLINEARIDADE	40
4	ELABORAÇÃO DO MODELO	41
4.1	ANÁLISE GRÁFICA DAS VARIÁVEIS	41
4.2	SIMULAÇÕES DAS REGRESSÕES ATRAVÉS DO MINITAB	48
4.2.1	Simulação com o conjunto de vinte e seis variáveis	49

4.2.2	Simulação com o conjunto de vinte variáveis	60
4.2.3	Simulação com o conjunto de quinze variáveis	66
4.2.4	Simulação com o conjunto de dez variáveis	72
4.2.5	Simulação de um modelo de regressão com o IBOVESPA deslocado	77
4.2.6	Simulação de um modelo de regressão com variáveis deslocadas.....	86
5	CONCLUSÃO	93
6	REFERÊNCIAS BIBLIOGRÁFICAS	94

1 INTRODUÇÃO

1.1 DEFINIÇÃO DO PROBLEMA

Aumentar o rendimento de aplicações financeiras é um pensamento recorrente de todos aqueles que possuem algum investimento, seja em Renda Fixa ou em Renda Variável. Este comportamento remete à ideia de como investir de forma inteligente e eficiente, a fim de se obter rendimentos maiores que aqueles provenientes de aplicações na Poupança. A princípio, podem ser encontradas algumas respostas triviais para esta questão, como a simples diversificação do capital investido, mas que para aqueles sem formação em finanças se torna muito complexa.

De acordo com (KIYOSAKI e LECHTER, 2000, p. 41), “Os pobres e a classe média trabalham pelo dinheiro. Os ricos fazem o dinheiro trabalhar por eles”, o que significa dizer que os ricos sabem como fazer suas aplicações renderem mais, ou melhor, possuem estratégias que tornam seus investimentos mais eficientes que a Poupança.

Em geral, pode-se mencionar duas classes de investimentos, Renda Fixa e Renda Variável, e a principal diferença entre elas se refere às oscilações do mercado, sendo a primeira menos sujeita a estas que a segunda.

Investir em ações de uma Bolsa de Valores é fazer um investimento em Renda Variável, e este está relacionado a rendimentos e riscos variáveis, não só pela volatilidade deste mercado mas pela complexidade em se formar boas carteiras de ações. Os obstáculos são inerentes a este tipo de investimento, ao mesmo tempo que a necessidade de se ter uma boa estratégia para otimizar a eficiência de uma carteira.

Existem diversos índices que servem como balizadores para os investidores em ações da Bolsa de Valores de São Paulo (BOVESPA) e o principal deles é o IBOVESPA, que é um índice influenciado pela saúde financeira das empresas que o compõem, e estas são influenciadas por notícias nacionais e internacionais no âmbito social, econômico ou político que afetam suas decisões.

Apesar de ser o principal índice da BOVESPA, o IBOVESPA não é muito utilizado por pequenos investidores, por ser muito difícil e arriscado utilizá-lo como direcionador das decisões de investimento, ou seja, não é possível utilizar o índice por si

só para prever movimentos do mercado acionário.

Então, para diminuir o risco gerado pelas incertezas do mercado, é usual construir carteiras bem diversificadas para que se possa maximizar os lucros e minimizar os riscos. Mas, supondo que fosse possível ter uma boa estimativa da tendência do IBOVESPA para os próximos dias, semanas ou meses, também seria possível estender esta estimativa aos ativos que compõem uma carteira e então garantir um melhor retorno, considerando uma perspectiva futura do mercado.

Para resolver a questão apresentada e tentar identificar como o IBOVESPA se comporta, neste trabalho é elaborado um estudo sobre a relação desse índice com o movimento histórico de diversas variáveis, que podem influenciar o mercado acionário brasileiro.

1.2 OBJETIVO DO TRABALHO

Uma vez idealizado o problema, este trabalho tem como objetivo mapear as principais variáveis que apresentam importância para influenciar o mercado de ações brasileiro e apresentar um estudo que mostre como estas estão relacionadas ao IBOVESPA e apresentar um modelo matemático para estimar os movimentos futuros do IBOVESPA, a fim de maximizar a eficiência de um portfólio de ações.

Com esse modelo, deseja-se alterar a percepção que se tem sobre a administração de um portfólio de ações, por exemplo, utilizando-o para determinar os pesos ótimos das ações de uma carteira, maximizando o ganho, e reduzindo o risco não só pela diminuição da variância mas também levando em consideração o nível de aderência do modelo à realidade, uma vez que se pode estimar o movimento do IBOVESPA e correlacioná-lo a qualquer ativo que o compõe de fato.

Espera-se, também, com esse estudo, colocar em prática conhecimentos adquiridos durante o curso de Engenharia Financeira e gerar oportunidade para novas ideias.

1.3 JUSTIFICATIVA DO TRABALHO

A principal hipótese é que existe uma relação entre diversas variáveis de âmbito econômico e o índice IBOVESPA. Portanto, pressupõe-se que este assunto é de interesse de instituições financeiras, tais como corretoras e bancos.

O estudo também possui importância para o setor acadêmico, uma vez que estão sendo mesclados assuntos distintos, o que pode ser proveitoso para outras áreas utilizar o modelo ou método pelo qual o mesmo foi construído. Nesse estudo específico estão sendo tratadas questões como modelos de regressão e análise estatística de dados históricos.

É possível enumerar vários trabalhos acadêmicos que mostram a relação entre os preços de ativos e variáveis macroeconômicas, tais como (SCHOR, BONOMO e PEREIRA, 2002), (NUNES, JR. e SEABRA, 2003), (GROPPÔ, 2004) e (NETO, 2006).

Segundo (GROPPÔ, 2004), a taxa básica de juros da economia tem impacto significativo no IBOVESPA, e portanto mostra a importância exercida pela taxa de juros na economia brasileira, sugerindo assim que os agentes econômicos que investem no mercado acionário brasileiro veem o investimento em renda fixa como sendo o grande substituto das aplicações em ações.

É possível verificar que o tema abordado possui certa influência sobre o cotidiano das pessoas e empresas, sendo que nesse caso, espera-se estimar valores futuros para o IBOVESPA, através do modelo matemático aqui desenvolvido, para apoiar decisões de investimento.

1.4 ESTRUTURA DO TRABALHO

O trabalho está assim estruturado. Além dessa primeira parte introdutória, o estudo contará com mais quatro capítulos. O segundo capítulo compreende uma revisão sobre as variáveis selecionadas como entrada do modelo de previsão bem como o período de coleta dos dados. No terceiro capítulo, são apresentados o referencial teórico e metodológico que sustentarão o modelo apresentado. No quarto capítulo é abordado o desenvolvimento do modelo matemático em si para estimação do IBOVESPA, o enfoque

na utilização desse modelo para balanceamento de uma carteira, além de discutir algumas comparações entre modelos de regressão, a fim de se verificar como se comportam em relação à realidade diante das variáveis selecionadas e por fim, no quinto e último capítulo, têm-se as conclusões.

1.5 LIMITAÇÕES DO TRABALHO

Por se tratar de um estudo empírico de um modelo matemático baseado em dados históricos, alguns obstáculos podem ser encontrados, tais como fontes de informações homogêneas, influência de eventos sazonais, políticos, sociais, econômicos, etc. Em resumo, a qualidade dos dados obtidos tem grande importância para o resultado deste estudo.

2 ESCOLHAS DAS VARIÁVEIS DE ENTRADA DO MODELO

Segundo (GUJARATI e PORTER, 2011, p. 39), a análise de regressão estuda a dependência de uma variável dependente em relação a uma ou mais variáveis independentes ou explanatórias, tendo como objetivo prever o valor médio da primeira em termos dos valores conhecidos das segundas.

Como já mencionado anteriormente, a variável dependente é o índice IBOVESPA. Agora se faz necessário propor as variáveis explanatórias que farão parte do modelo. Baseado em artigos, teses e livros com estudos semelhantes a este, onde o autor visa explicar a correlação ou relação causal entre variáveis macroeconômicas e o índice IBOVESPA, são selecionadas as variáveis que apresentam maior significância e que possuam séries históricas disponíveis para consulta pública.

Em seu estudo para proposição de um modelo de regressão para o IBOVESPA, (NETO, 2006) estabeleceu cinco grupos principais de variáveis, que julgou estarem correlacionados ao IBOVESPA, sendo eles índices de ações, moedas, índices da dívida pública brasileira, *commodities* e taxa básica de juros.

(GROPPÔ, 2004), por sua vez, também utiliza a taxa básica de juros e acrescenta que o preço do petróleo no mercado internacional também tem um importante papel na explicação dos movimentos do índice da Bolsa de Valores de São Paulo.

Alguns outros trabalhos também foram verificados a fim de estabelecer um grupo consistente de variáveis para o estudo, tais como (PIMENTA e HIGUCHI, 2008), (SILVA e CORONEL, 2012) e (GAIO, AMBROZINI, *et al.*, 2014).

Com base nos estudos citados, os seguintes grupos de variáveis foram selecionados para a elaboração do modelo proposto neste trabalho, não significando, necessariamente, que todas as variáveis serão utilizadas:

- ✓ Índices de Ações;
- ✓ Moedas;
- ✓ Medida Risco Brasil;
- ✓ *Commodities*;
- ✓ Taxa de Juros Futuros.

2.1 PERÍODO E FONTES PARA COLETA DOS DADOS

Como existe uma variedade de dados coletados e nem sempre os mesmos são encontrados na mesma fonte, convencionou usar-se um período de dois anos para trás a partir da data de coleta das informações. Portanto o período de coleta dos dados é de 24/01/2013 a 23/01/2015.

Os dados foram coletados em *sites* de Bolsas de Valores, como o IBOVESPA, em *sites* de notícias econômicas e financeiras, como o INVESTING, o UOL e o INFOMONEY, e o site do IPEA – Instituto de Pesquisa Econômica Aplicada.

2.2 TRATAMENTO E PREPARAÇÃO DOS DADOS

Durante a coleta de dados, foi verificado que para algumas variáveis, em certos momentos das séries, não havia dados de acordo com as instituições, muitas vezes devido aos feriados ou algum outro evento que fizesse o índice perder liquidez naquela data específica.

Para essas datas, convencionou-se usar a média aritmética entre os períodos anterior e posterior para preencher o dado faltante.

A partir de alguns trabalhos estudados (MEDEIROS, 2006) e (NETO, 2006), e devido aos bons resultados apresentados nestes, decidiu-se utilizar as ferramentas Minitab versão 17 e o Microsoft Excel versão 2013 para os cálculos das regressões e análises estatísticas.

2.3 ÍNDICES DE AÇÕES

Os índices de Ações são indicadores que normalmente englobam as principais ações negociadas na Bolsa de Valores a que se referem. Cada índice possui sua própria metodologia de cálculo, mas basicamente são médias ponderadas por um determinado fator, por exemplo o volume de negociações.

Normalmente as Bolsas de Valores apresentam um ou mais índices formados por ações de diversos setores ou por outros critérios de âmbito econômico. Assim, esses

índices refletem como a economia está se desenvolvendo nos setores a que estes se referem.

De acordo com (NETO, 2006, p. 30), quando um país cresce diversos índices de ações refletem esse movimento, mostrando que eles estão intimamente ligados à economia.

Com base nos trabalhos (NETO, 2006) e (MEDEIROS, 2006), e pesquisas em sites de notícias econômicas foram selecionados os índices de Ações apresentados na Tabela 1.

Tabela 1 – Índices de ações selecionados para o estudo do modelo de previsão do IBOVESPA.

ÍNDICE DE AÇÕES	PAÍS DE ORIGEM
IBOVESPA	BRASIL
CAC40	FRANÇA
DAX30	ALEMANHA
DJI	EUA
FTSEMIB	ITÁLIA
HSENG	CHINA
IBEX35	ESPAÑHA
IPCMEX	MÉXICO
MERVAL	ARGENTINA
NASDAQC	EUA
NIKKEI225	JAPÃO
SHNGYC	CHINA
SMI	SUIÇA
SP500	EUA

A seguir são dadas breves descrições sobre cada índice de ações selecionado para compor o modelo de previsão do IBOVESPA.

2.3.1 CAC40

O CAC40 é o principal índice de ações da França e é um dos mais negociados pelos investidores franceses por ser um dos mais importantes do mercado EURONEXT. O índice é calculado a partir dos resultados das ações das quarenta empresas francesas de maior desempenho e estão entre as cem maiores empresas listadas na Bolsa de Valores Francesa.

2.3.2 DAX30

O DAX30 é o índice da Bolsa de Valores alemã e é equivalente ao CAC40 francês. É considerado o principal índice da Bolsa de Valores de Frankfurt e uma das suas principais características é o volume de transações que ele gera, tornando-o o índice europeu mais negociado do mundo.

Como o nome sugere, o índice é composto pelas trinta empresas alemãs mais cotadas na Bolsa de Valores de Frankfurt, e tem como objetivo mostrar a imagem econômica da Alemanha, incluindo todos os setores de atividades.

2.3.3 DJI

O Dow Jones é o índice de ações que representa a economia americana, e foi o primeiro índice a ser cotado na Bolsa de Valores de Nova Iorque. É também o índice mais antigo do mercado acionário no âmbito internacional.

O índice é formado por apenas trinta empresas, consideradas como as mais rentáveis e mais representativas do setor industrial nos Estados Unidos. Exemplos de empresas que fazem parte desse índice são Coca-Cola e Intel.

2.3.4 FTSEMIB

O FTSEMIB é o índice acionário que representa a Bolsa de Valores italiana, que substituiu o índice anterior MIB30 em setembro de 2004. O índice foi administrado pela Standard & Poors desde o seu início até junho de 2009, quando essa responsabilidade foi passada para o *FTSE Group*. O índice representa as quarenta ações italianas com maior liquidez.

2.3.5 HSENG

O HSENG é o principal índice de desempenho da Bolsa de Valores de Hong Kong. Ele é composto pelas cinquenta ações de maior liquidez, que representam cerca

de 58% da capitalização da Bolsa de Valores chinesa.

2.3.6 IBEX35

O índice IBEX35 é composto pelas 35 ações mais líquidas negociadas na Bolsa de Valores espanhola, e é utilizado como referência nacional e internacional na contratação de derivativos. Tecnicamente é um índice de preços, ponderado pela capitalização e ajustado pelo *free float*¹ de cada empresa membro do índice.

2.3.7 IPCMEX

O índice mexicano IPCMEX é um índice de capitalização ponderada das principais ações negociadas no mercado acionário mexicano. É um índice composto pelas trinta e cinco ações mais negociadas na Bolsa de Valores mexicana.

2.3.8 Merval

O Merval é o principal índice da bolsa de valores da Argentina e o terceiro mais importante da América Latina atrás apenas da BM&FBOVESPA em São Paulo e da Bolsa de Valores mexicana na Cidade do México. Representa as ações mais negociadas do mercado argentino. Durante 2013, o Merval foi o segundo mercado de ações entre os mais rentáveis do mundo.

2.3.9 NASDAQC

NASDAQ é um mercado de ações automatizado norte-americano onde estão listadas mais de 2800 ações de diferentes empresas, em sua maioria de pequena e média capitalização.

¹ "Free float é a quantidade percentual de ações livres à negociação no mercado. São aquelas ações que não pertencem a acionistas estratégicos, como: controladores e diretores da companhia e acionistas que detenham mais de 5% do capital total da empresa[...]" (INVESTPEDIA).

O índice NASDAQC caracteriza-se por reunir empresas de alta tecnologia em eletrônica, informática, telecomunicações, biotecnologia, etc.

2.3.10 NIKKEI225

O NIKKEI225 é um índice da Bolsa de Valores de Tokio, e é calculado diariamente pelo jornal Nihon Keizai Shimbun (Nikkei) desde 1950. É um índice ponderado pelos preços, e os ativos são revistos uma vez por ano. Atualmente, o Nikkei é composto pelas ações mais negociadas no mercado japonês.

2.3.11 SHNGYC

O SHNGYC é um índice da Bolsa de Valores de Xangai que representa todas as ações negociadas no mercado acionário chinês.

2.3.12 SMI

O SMI é o índice da Bolsa de Valores suíça. Ele lista as 20 maiores empresas suíças com maiores volumes de negociação que constam na *SIX Swiss Exchange*, a bolsa de valores de Zurique.

2.3.13 SP500

O SP500 é um índice americano composto por quinhentas ações qualificadas devido ao seu tamanho de mercado, sua liquidez e sua representação industrial.

SP refere-se a uma empresa chamada Standard & Poors, que se dedica principalmente à consultoria financeira, entre outras atividades de âmbito econômico e financeiro.

2.4 MOEDAS

A moeda é uma das principais variáveis econômicas e sua valorização em relação a outras moedas mostra o nível de confiança que a população tem nas políticas monetárias e fiscais de seus países, e é o preço das moedas que viabiliza a troca entre elas próprias e estabelece a quantidade que uma moeda pode ser adquirida por unidade de outra moeda.

A valorização da moeda em relação a outras é de grande importância para exportadores e importadores, que realizam suas transações através do câmbio, e ela pode determinar o nível de atividade econômica das empresas, refletindo diretamente no mercado acionário, por exemplo, através de uma crise cambial.

Da mesma forma como os índices de ações foram selecionados, as moedas indicadas na Tabela 2 foram selecionadas com base nos trabalhos (NETO, 2006) e (MEDEIROS, 2006).

Tabela 2 – Moedas selecionadas para o estudo do modelo de previsão do IBOVESPA.

MOEDA	PAÍS
DÓLAR AMERICANO	EUA
EURO	ZONA DO EURO
LIBRA ESTERLINA	REINO UNIDO
PESO ARGENTINO	ARGENTINA
IENE	JAPÃO

2.5 MEDIDA RISCO BRASIL

O risco país é uma medida criada pelo banco norte-americano JP Morgan, com o intuito de medir a percepção de risco dos investidores em relação a diversos países, tomando como base a cotação dos ativos da dívida externa destes países negociados no mercado internacional.

A partir do *spread*, ou diferencial de juros que estes títulos pagam em relação aos títulos do tesouro norte-americano de prazo semelhante, o banco calcula a medida de risco ponderada. Assim, se o diferencial médio é de 10%, o risco país é de 1.000 pontos base, medida criada para capturar pequenas mudanças no *spread*.

Quanto maior o risco país, maior é o prêmio que deve ser pago aos investidores para compensá-los de assumir esse risco. Assim, quando o risco Brasil é igual a 260 pontos, significa que a carteira teórica de papéis emitidos pelo país paga 2,60% a mais que uma carteira teórica de papéis emitidos pelo EUA de mesmo prazo.

Pode-se dizer que o risco Brasil mostra o nível de insegurança de investidores estrangeiros aplicar o seu capital no país e segundo o Banco Central, o aumento persistente do risco Brasil piora as condições de acesso ao financiamento externo e contribui para a desvalorização do Real.

Como muitas empresas que compõem o IBOVESPA reagem à desvalorização do Real, essa variável é incluída no estudo do modelo de previsão do índice.

2.6 COMMODITIES

As *commodities* são produtos cultivados ou de extração mineral, que podem ser estocados por certo tempo sem perda sensível de suas qualidades, como suco de laranja congelado, soja, trigo, bauxita, prata ou ouro.

As *commodities* são uma forma de investimento, uma opção entre as tantas opções de investimento no mercado, como poupança ou fundos de investimento.

Algumas empresas exportadoras de *commodities* possuem um peso grande no IBOVESPA. Assim, quando o preço das *commodities* sobe, o valor das ações das empresas exportadoras sobem e consequentemente o índice IBOVESPA sobe.

A Tabela 3 mostra as *commodities* que foram selecionadas para compor o modelo de previsão do IBOVESPA.

Tabela 3 – Commodities selecionadas para o estudo do modelo de previsão do IBOVESPA.

COMMODITIES
OURO
SOJA
CAFÉ
PETRÓLEO
TRIGO
AÇÚCAR NY

2.7 TAXA DE JUROS FUTUROS

A taxa básica de juros (SELIC no caso do Brasil) é um instrumento utilizado pelo Banco Central para controlar a inflação no país e está ligada diretamente ao ritmo da economia no país.

Se a SELIC cai, a população tem maior acesso ao crédito, o que leva a um consumo maior, e esta demanda maior pode pressionar os preços caso a indústria não esteja preparada para atender a este consumo maior. Com o aumento dos preços a inflação aumenta. Por outro lado, se a SELIC sobe, o consumo e investimentos diminuem, a economia desacelera e evita-se que os preços subam, ou seja, a inflação reduz.

O mercado de juros futuros é um dos mais sensíveis às decisões do COPOM de aumentar ou reduzir a SELIC, pois interfere diretamente nos negócios. Assim, a taxa de juros futuros é um importante fator para saber qual a expectativa do mercado em relação à SELIC e, por conseguinte, em relação às outras variáveis macroeconômicas, principalmente a inflação.

3 REFERENCIAL TEÓRICO

Nesse capítulo são apresentados alguns conceitos importantes para o entendimento do processo de análise de séries temporais financeiras, e com base no conhecimento adquirido durante o curso, propor um modelo matemático para prever valores do IBOVESPA, levando em consideração algumas variáveis macroeconômicas.

Em princípio, não haveria diferenças entre a análise de séries temporais financeiras e aquelas que ocorrem em outras áreas, como oceanografia, meteorologia, engenharia, etc. De fato, a maioria das técnicas de análise de séries temporais são aplicáveis em diversas áreas. Porém, uma característica presente em séries temporais financeiras é o que se convencionou chamar de *volatilidade*, que pode ser definida de várias maneiras, mas não é diretamente observável (MORETTIN, 2011).

Uma dessas definições diz que a volatilidade é uma medida de dispersão dos retornos de um ativo do mercado financeiro, dentre as quais pode-se citar a medida de variância, que é a mais utilizada para verificar a dispersão de uma série temporal financeira.

Segundo (GUJARATI e PORTER, 2011), seja X uma variável aleatória e seja $E(X) = \mu$. A distribuição, ou dispersão, dos valores de X em torno de um valor esperado pode ser mensurada pela variância, definida como

$$var(X) = \sigma^2 = E(X - \mu)^2 \quad (3-1)$$

A variância indica quão próximos ou distantes os valores individuais de X estão distribuídos em torno de seu valor médio.

3.1 TIPOS DE DADOS

Em séries temporais financeiras tem-se que levar em consideração os tipos de dados que são analisados. Pode-se considerar observações igualmente espaçadas e consecutivas, como por exemplo, observações diárias, semanais, mensais, etc. Nesse estudo é utilizado o intervalo diário e o preço de fechamento dos ativos selecionados. Na

Figura 1 pode ser observada a série diária do índice IBOVESPA do ano de 2014 para melhor ilustração. Os dados ilustram a evolução do valor de fechamento do índice ao longo do ano.

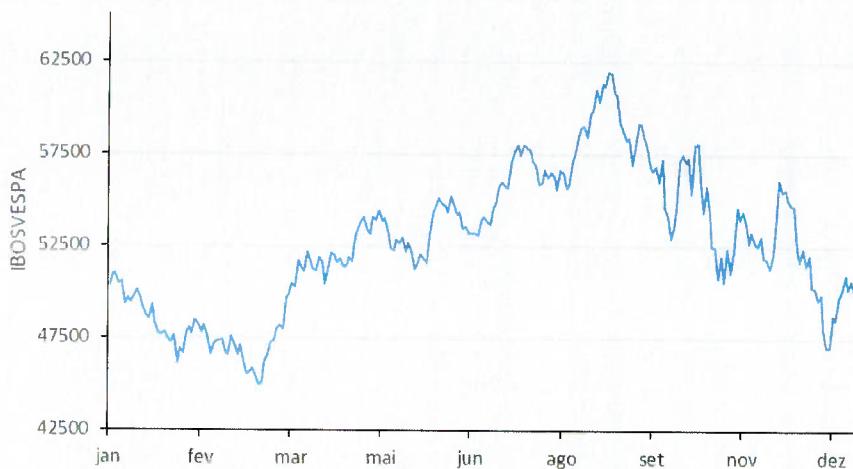


Figura 1 – Série histórica das cotações do IBOVESPA em 2014 (UOL).

3.2 RETORNOS

Volatilidade também é uma medida de risco muito usada em mercados acionários no âmbito nacional e internacional, e segundo (MORETTIN, 2011, p. 7), um dos objetivos em finanças é determinar o risco que uma carteira de ativos pode apresentar em termos de variações de preços dos ativos, ou seja, através da variação dos retornos.

Pode-se medir o *retorno líquido* de um ativo entre dois instantes, $t - 1$ e t , através da seguinte fórmula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}} \quad (3-2)$$

em que:

R_t é denotado em percentagem e também é chamado de *taxa de retorno*;

P_t é o preço do ativo no instante t ;

P_{t-1} é o preço do ativo no instante $t - 1$.

Também pode-se utilizar o *retorno composto continuamente*, ou simplesmente *log-retorno* como:

$$r_t = \ln \frac{P_t}{P_{t-1}} \quad (3-3)$$

Segundo (MORETTIN, 2011, p. 8), é preferível trabalhar com retornos, que são livres de escala e têm propriedades estatísticas mais interessantes como estacionariedade e ergocidade.

A Figura 2 representa a série temporal dos retornos do índice IBOVESPA ao longo do ano de 2014.

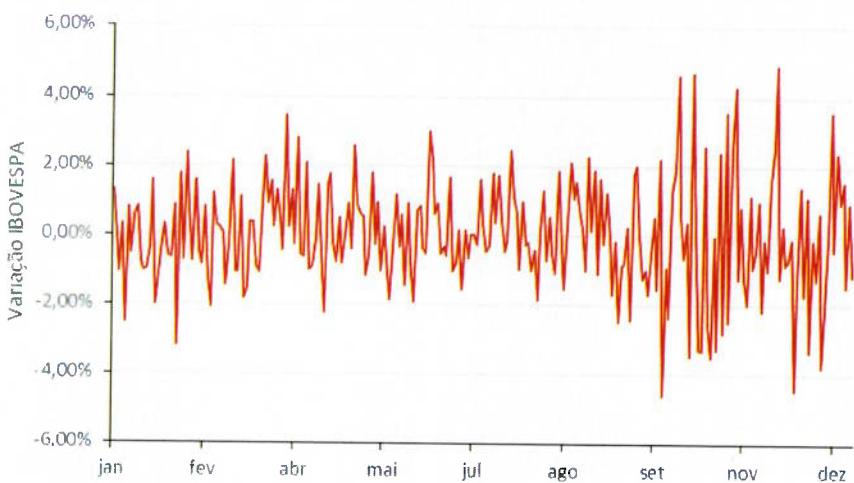


Figura 2 – Série histórica dos retornos do índice IBOVESPA em 2014 (UOL).

O principal objetivo desse estudo é apresentar como algumas variáveis financeiras influenciam o IBOVESPA através de um modelo de regressão clássica, e com o intuito de enriquecer o estudo, também é apresentado um modelo para série temporal do tipo ARMA, utilizado, neste estudo, para modelar os resíduos gerados pelos outros modelos. Com isso, espera-se aumentar a qualidade das previsões do índice IBOVESPA.

3.3 PROCESSOS ESTOCÁSTICOS

Avançando um pouco nas definições, e de acordo com (MORETTIN e TOLOI, 2006, p. 19), os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, uma conjunto de variáveis aleatórias que representam a evolução temporal de algum sistema.

Pode-se considerar um número muito grande de modelos diferentes para descrever o comportamento de uma série particular. A construção destes modelos depende de vários fatores, tais como o comportamento do fenômeno ou o conhecimento *a priori* que se tem de sua natureza e do objetivo da análise (MEDEIROS, 2006, p. 19).

Seja T um conjunto arbitrário. Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.

Nestas condições, um processo estocástico é uma família de variáveis aleatórias, que supõe-se definidas em um mesmo espaço de probabilidade (Ω, A, P) , onde Ω representa o espaço amostral, A o conjunto de eventos e P como uma medida positiva em σ -álgebra, tal que $P(\Omega) = 1$.

A Figura 3 ilustra um processo estocástico, onde para cada $t \in T$, tem-se uma variável aleatória $Z(t, \omega)$, com uma distribuição de probabilidade, e é possível que a função de densidade de probabilidade no instante t_2 seja diferente do instante t_1 , para dois instantes quaisquer, mas a situação usual é aquela em que a função densidade de probabilidade de $Z(t, \omega)$ é a mesma, para todo $t \in T$.

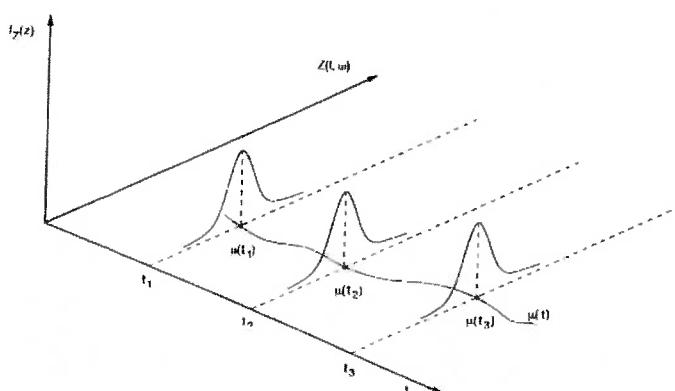


Figura 3 – Um processo estocástico interpretado como uma família de variáveis aleatórias (MORETTIN e TOLOI, 2006, p. 20).

3.4 FUNÇÃO DE DENSIDADE DE PROBABILIDADE E DISTRIBUIÇÃO DE PROBABILIDADE

Seja X uma variável aleatória discreta que tem valores distintos de $x_1, x_2, \dots, x_n, \dots$. Então, a função:

$$f(x) = \begin{cases} P(X = x_i), & i = 1, 2, \dots, n, \dots \\ 0, & x \neq x_i \end{cases}$$

é chamada função de densidade de probabilidade discreta (FDP) de X , em que $P(X = x_i)$ significa a probabilidade de que a variável aleatória X tenha o valor de x_i .

Segundo (GUJARATI e PORTER, 2011, p. 804), uma distribuição de probabilidade pode ser explicada por algumas poucas características. Duas das características mais utilizadas são a média e a variância.

3.5 REGRESSÃO LINEAR

O estudo da regressão tenta estimar ou prever o valor médio $E(Y)$ com base em valores conhecidos de uma ou mais variáveis explanatórias.

A análise de regressão diz respeito ao estudo da dependência de uma variável dependente em relação a uma ou mais variáveis explanatórias visando prever o valor médio (populacional) da primeira em termos dos valores conhecidos ou fixados (em amostragens) das segundas.

“Nas relações estatísticas entre variáveis, lida-se essencialmente com variáveis aleatórias ou estocásticas, isto é, variáveis que têm distribuições probabilísticas” (GUJARATI e PORTER, 2011, p. 42).

É importante notar que uma relação estatística por si só não implica logicamente em uma relação de causa e efeito. Ou seja, uma variável independente pode apresentar forte correlação com uma variável dependente em um determinado período, mas não necessariamente o valor de uma foi causado pelo valor da outra. Seria necessário citar alguma teoria para justificar a relação entre elas.

3.6 MODELO DE REGRESSÃO COM DUAS VARIÁVEIS

A análise de regressão linear simples, consiste em uma análise estatística para determinar a relação entre uma variável dependente e uma variável explanatória. Ou seja, deseja-se obter uma equação matemática que explica como uma variável se comporta em relação à outra.

Para tentar estabelecer uma equação que explique esse fenômeno, pode-se fazer uso de um gráfico de dispersão e verificar como se comportam os valores da variável dependente (Y) em relação à variável explanatória (X).

O comportamento de uma variável em relação a outra pode se apresentar em vários formatos: linear, quadrático, cúbico, etc. Mas para chegar ao modelo que melhor representa a relação entre as variáveis, é necessário entender que tipo de curva melhor se aproxima dos pontos apresentados no gráfico de dispersão.

A Figura 4 ilustra esse tipo de gráfico e a relação entre o IBOVESPA e o Dólar Americano no período de 24/01/2013 a 23/07/2014 e a aproximação da relação entre as variáveis através de uma reta, conhecida como linha de regressão populacional (LRP). O termo “populacional” se refere ao fato que, neste exemplo, se está considerando todos os dados entre o período mencionado. Na realidade, se considerado um período maior, os dados selecionados para representar o gráfico seriam uma “amostra” da população.

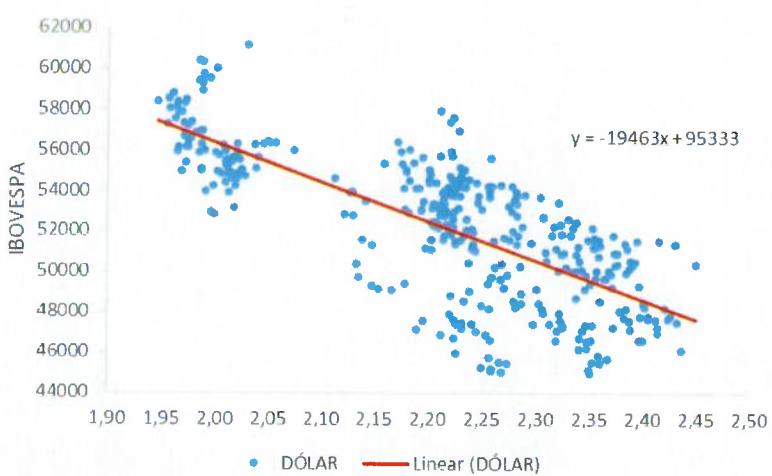


Figura 4 – Gráfico de dispersão IBOVESPA x Dólar americano.

Na Figura 4 é possível verificar que a maioria dos pontos não se ajustam à reta do modelo proposto. Isso acontece devido ao fenômeno que está em estudo, não ser um fenômeno matemático e sim um fenômeno que está sujeito a influências que acontecem ao acaso. Neste exemplo a linha de regressão populacional, LRP, é uma reta, mas poderia ser uma curva, que o conceito seria o mesmo.

A equação apresentada na Figura 4 se baseia em uma função de regressão de 1º grau representada pela equação

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (3-4)$$

em que

Y_i = valor da variável Y no i -ésimo valor da variável X ,

β_1 = constante da regressão,

β_2 = coeficiente da regressão,

X_i = i -ésimo valor da variável X ,

u_i = desvio individual de Y_i em torno de seu valor esperado, ou técnicamente distúrbio estocástico ou resíduos.

Segundo (GUJARATI e PORTER, 2011), a função de regressão populacional não pode ser observada diretamente e é necessário estimá-la por meio de uma função de regressão amostral (FRA) representada pela equação:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (3-5)$$

Em termos da função de regressão amostral, o Y_i observado pode ser expresso como:

$$Y_i = \hat{Y}_i + \hat{u}_i \quad (3-6)$$

em que \hat{Y}_i é o valor estimado de Y_i através de uma amostra dos dados. E a partir de (3-

5) e (3-6), tem-se que:

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \quad (3-7)$$

onde os resíduos são simplesmente a diferença entre os valores observados e estimados de Y .

Uma técnica comumente utilizada para estimar a FRA é o método dos Mínimos Quadrados Ordinários (MQO), cujo objetivo é encontrar os valores de β_1 e β_2 que apresentam a menor soma dos quadrados dos resíduos.

Elevando-se ambos os membros da equação (3-7) ao quadrado e aplicando-se o somatório, tem-se:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \quad (3-8)$$

Para encontrar o mínimo para uma equação, deve-se derivá-la em relação à variável de interesse e igualá-la a zero. Derivando parcialmente a equação (3-8) em relação a $\hat{\beta}_1$ e $\hat{\beta}_2$, e igualando a zero, depois de simplificações e manipulações algébricas, obtêm-se os estimadores dados pelas equações (3-9) e (3-10):

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \quad (3-9)$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad (3-10)$$

em que \bar{X} e \bar{Y} são medidas amostrais de X e de Y e onde se define $x_i = (X_i - \bar{X})$ e $y_i = (Y_i - \bar{Y})$.

Uma vez obtidas as estimativas de $\hat{\beta}_1$ e $\hat{\beta}_2$, pode-se escrever a FRA como:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (3-11)$$

3.7 COEFICIENTE DE DETERMINAÇÃO R^2

De acordo com (GUJARATI e PORTER, 2011, p. 97), o valor de R^2 é conhecido como o coeficiente de determinação e é o indicador mais usado para medir a qualidade do ajuste de uma linha de regressão, ou seja, R^2 mede o percentual da variação total de Y explicada pelo modelo de regressão.

O valor de R^2 é dado pela equação:

$$R^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2} \quad (3-12)$$

3.8 REGRESSÃO MÚLTIPLA

Pode-se dizer que o caso de regressão múltipla é uma extensão dos conceitos do caso da regressão de duas variáveis.

O modelo de regressão simples, com duas variáveis, nem sempre descreve da melhor forma a relação entre a variável dependente e a variável independente, uma vez que podem existir outras variáveis, que possivelmente afetam o comportamento da variável dependente, que não foram consideradas na elaboração do modelo.

Segundo (GUJARATI e PORTER, 2011), o modelo de regressão múltipla mais simples possível é o de três variáveis, uma dependente e duas explanatórias e generalizando a função de regressão simples com duas variáveis, pode-se escrever a função de regressão com três variáveis do seguinte modo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (3-13)$$

em que Y_i é a variável dependente, X_2 e X_3 , as variáveis explanatórias, u o termo residual e i , o indicador da i -ésima observação. Na equação (3-13), β_1 é a constante da regressão. Os coeficientes β_2 e β_3 são denominados coeficientes parciais de regressão.

Da mesma forma que para a Regressão Linear Simples, é utilizado o método dos Mínimos Quadrados Ordinários (MQO), partindo-se da função de regressão amostral correspondente à função de regressão populacional:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \quad (3-14)$$

em que \hat{u}_i é o termo residual, a contrapartida amostral do termo de erro estocástico u_i . Conforme explicado anteriormente, o cálculo do MQO consiste na escolha dos coeficientes β_1 , β_2 e β_3 que minimizam a soma dos quadrados dos resíduos $\sum \hat{u}_i^2$, representado pela equação:

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \quad (3-15)$$

O método para encontrar os coeficientes segue os mesmos procedimentos da Regressão Linear Simples, e após simplificações e manipulações algébricas, obtêm-se os coeficientes (3-16), (3-17) e (3-18):

$$\hat{\beta}_1 = Y_i - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \quad (3-16)$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad (3-17)$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad (3-18)$$

Uma vez obtidas as estimativas de β_1 , β_2 e β_3 , pode-se escrever a FRA como:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \quad (3-19)$$

Se uma série temporal é não estacionária, podemos estudar seu comportamento apenas pelo período de tempo em consideração. Cada conjunto de dados de série temporal, portanto, é específico a cada período. Como consequência, não é possível generalizá-lo para outros períodos. Sendo assim, para o propósito de previsão, o modelo pode ser de pouco valor prático.

3.9 MODELOS AUTO REGRESSIVOS

Um modelo auto regressivo (AR) é um caso especial dos modelos de filtro linear.

Para entender como os parâmetros de um modelo auto regressivo funcionam, considera-se um modelo Box Jenkins que contenha apenas um parâmetro AR. Este modelo pode ser escrito da seguinte forma:

$$Y_t = \alpha_1 Y_{t-1} + u_t \quad (3-20)$$

onde Y_t é uma série estacionária. O termo $\alpha_1 Y_{t-1}$ representa o ajuste do valor da série Y_t e α_1 é chamado parâmetro AR de ordem 1. O termo u_t representa um erro aleatório assumido no dado no período t .

O modelo indica que um valor Y_t na série é diretamente proporcional ao valor anterior mais um erro aleatório. Ou seja, o que acontece em um período depende somente do que aconteceu no período anterior, mais uma parcela de erro aleatório.

Isso não significa que o valor de Y_t na série não possa estar relacionado a mais de um termo passado, o que é possível e que pode ser descrito por uma fórmula genérica dos modelos auto regressivos:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + u_t \quad (3-21)$$

3.10 MODELOS DE MÉDIAS MÓVEIS

Um outro caso de filtro linear é o modelo de médias móveis (MA).

Embora os modelos MA sejam parecidos com os modelos AR, o conceito para o uso desse tipo de modelo é um tanto diferente. Os parâmetros do modelo de médias móveis estão relacionados apenas com os erros aleatórios que ocorrem nos períodos anteriores. Por exemplo, um modelo Box Jenkins com apenas parâmetros MA pode ser escrito de acordo com a equação genérica:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q} \quad (3-22)$$

em que μ é uma constante e u é o termo de erro residual. Aqui, Y_t é um processo de média móvel MA(q). Segundo (GUJARATI e PORTER, 2011), um processo de média móvel é apenas uma combinação linear de termos de erro de ruído branco.

3.11 MODELOS ARMA

O modelo misto combina os dois modelos apresentados anteriormente, ou seja, com parâmetros AR e MA. Modelos contendo ambos os tipos de parâmetros são chamados modelos ARMA e escritos na forma que se segue na equação (3-23):

$$Y_t = \theta + (\alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p}) + (\beta_0 u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q}) \quad (3-23)$$

onde θ representa um termo constante.

Em geral, em um processo ARMA (p, q), haverá termos auto regressivos p e termos de média móvel q .

3.12 IDENTIFICAÇÃO DE MODELOS ARMA

As principais ferramentas na identificação de modelos ARMA são a função de autocorrelação e a função de autocorrelação parcial.

3.12.1 Autocorrelação

A função de autocorrelação indica como uma série temporal se relaciona com ela mesma durante um intervalo de tempo. O número de períodos de intervalo normalmente é chamado de defasagem. Por exemplo, uma função de autocorrelação de defasagem dois mede como os valores de uma série estão correlacionados com os valores de dois períodos anteriores.

Um valor de autocorrelação sempre é um número entre -1 e $+1$. O valor $+1$

indica forte autocorrelação positiva, enquanto -1 significa o oposto, forte autocorrelação negativa. Uma autocorrelação próxima de zero indica que os dados não estão autocorrelacionados.

3.12.2 Autocorrelação Parcial

Autocorrelações parciais são outras medidas estatísticas que são usadas para estudar os relacionamentos entre os valores da série. Tratam-se de medidas complementares às autocorrelações, e a análise de seus padrões auxiliam na definição dos parâmetros dos modelos de médias móveis e auto regressivos. Os valores das autocorrelações parciais estão entre -1 e $+1$.

3.12.3 Escolha do Modelo AR(1), MA(1) e ARMA(1,1)

3.12.3.1 *Modelo AR(1)*

A função de autocorrelação apresenta padrão de decaimento exponencial e a função de autocorrelação parcial tem pico na defasagem *um* e truncamento após.

3.12.3.2 *Modelo MA(1)*

A função de autocorrelação apresenta pico na defasagem *um* e truncamento após, e a função de autocorrelação parcial decai em forma exponencial.

3.12.3.3 *Modelo ARMA(1,1)*

A função de autocorrelação apresenta padrão de decaimento e a função de autocorrelação parcial, também. Nesse caso, o modelo deve ter uma parte $AR(p)$ e uma parte $MA(q)$.

3.13 MULTICOLINEARIDADE

Em regressões múltiplas, quando se tem mais de duas variáveis explanatórias, é importante verificar se essas variáveis são correlacionadas. Se não houver nenhuma relação entre elas, pode-se dizer que são ortogonais.

Geralmente é difícil que as variáveis explanatórias sejam ortogonais, mas felizmente é possível encontrar variáveis pouco correlacionadas, o que não interfere no resultado da regressão. Mas se as variáveis forem muito correlacionadas, o modelo de regressão pode apresentar resultados pouco confiáveis.

Algumas características do modelo podem indicar problema de colinearidade, tais como:

- ✓ Coeficiente com sinal oposto ao esperado;
- ✓ Grandes mudanças dos coeficientes quando adicionada ou excluída alguma variável;
- ✓ Coeficiente não significante para uma variável teoricamente importante.

A multicolinearidade é um problema no ajuste do modelo que pode causar impactos na estimativa dos parâmetros. No Minitab, é apresentado o parâmetro VIF (*Variance Inflation Factor*) que serve para diagnosticar a multicolinearidade.

4 ELABORAÇÃO DO MODELO

Algumas diretrizes foram seguidas a fim de se elaborar o modelo de regressão múltipla:

- i. Usar estudos anteriores para selecionar o melhor conjunto de variáveis;
- ii. Manter o modelo o mais simples possível;
- iii. Utilizar variáveis que apresentem correlação forte, positiva ou negativa, com o IBOVESPA.

Apesar de terem sido selecionadas quase trinta variáveis, nem todas elas serão utilizadas no modelo, o que vai de encontro à segunda diretriz estabelecida no parágrafo anterior. Logo, para realizar uma triagem das variáveis que serão utilizadas, foram elaborados gráficos que apresentam a proximidade da variável independente com o IBOVESPA e possibilitam uma análise visual, a fim de identificar aquelas que apresentam comportamento similar ao índice, conforme apresentado nas Figuras 5 a 30.

4.1 ANÁLISE GRÁFICA DAS VARIÁVEIS

Nos gráficos de evolução dos índices, leia-se o eixo Y , primário, como o IBOVESPA e o secundário como o outro índice indicado na legenda, e no eixo X , leia-se o número da observação. Para o gráfico de dispersão, leia-se o eixo Y como o IBOVESPA e o eixo X como o outro índice indicado na legenda.



Figura 5 – Gráficos das séries temporais IBOVESPA, CAC40 e dispersão IBOVESPA x CAC40.

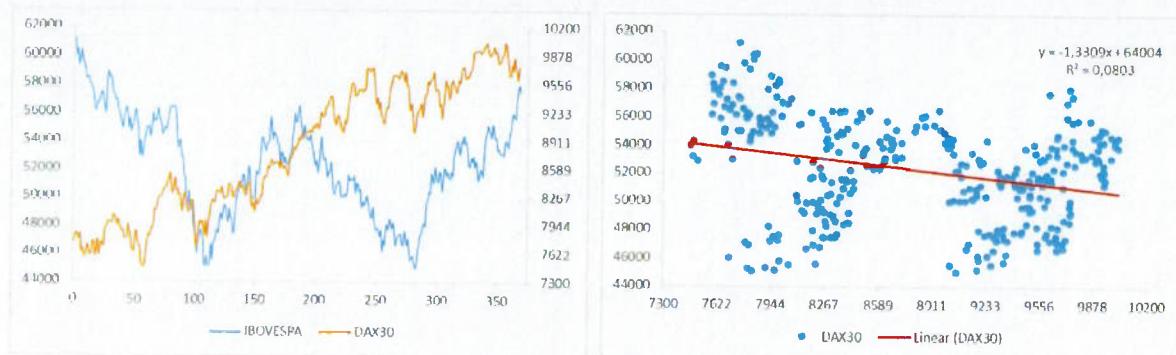


Figura 6 – Gráficos das séries temporais IBOVESPA, DAX30 e dispersão IBOVESPA x DAX30.

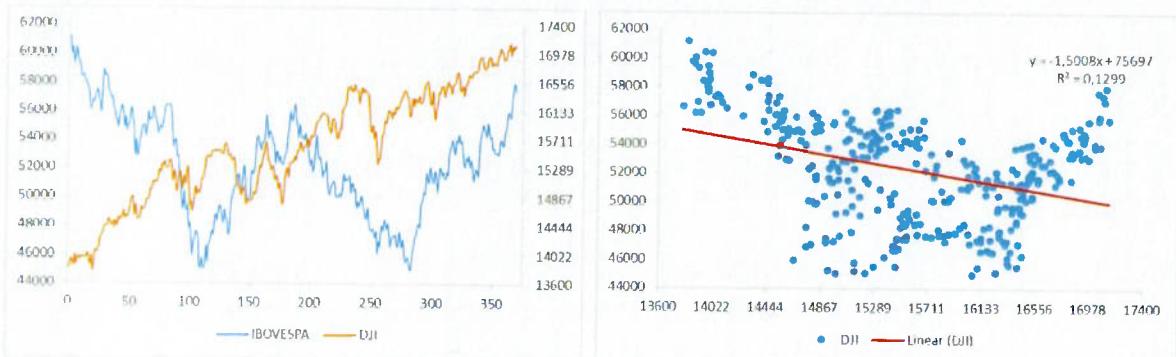


Figura 7 – Gráficos das séries temporais IBOVESPA, DJI e dispersão IBOVESPA x DJI.

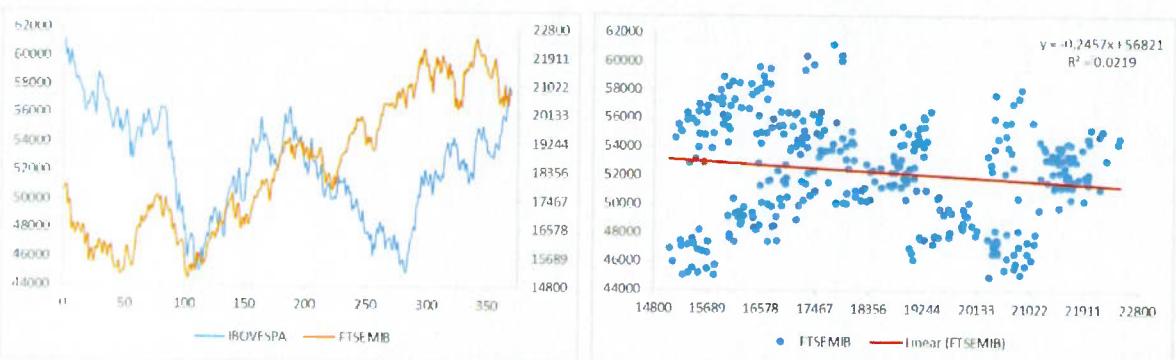


Figura 8 – Gráficos das séries temporais IBOVESPA, FTSEMIB e dispersão IBOVESPA x FTSEMIB.

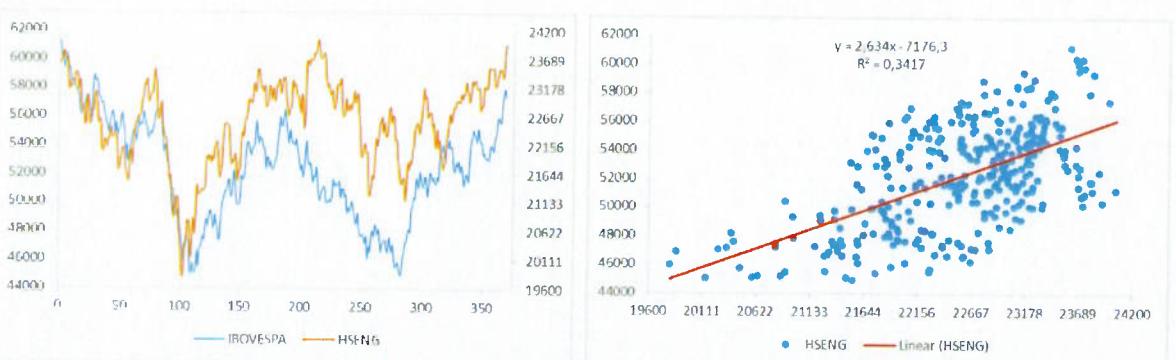


Figura 9 – Gráficos das séries temporais IBOVESPA, HSENG e dispersão IBOVESPA x HSENG.

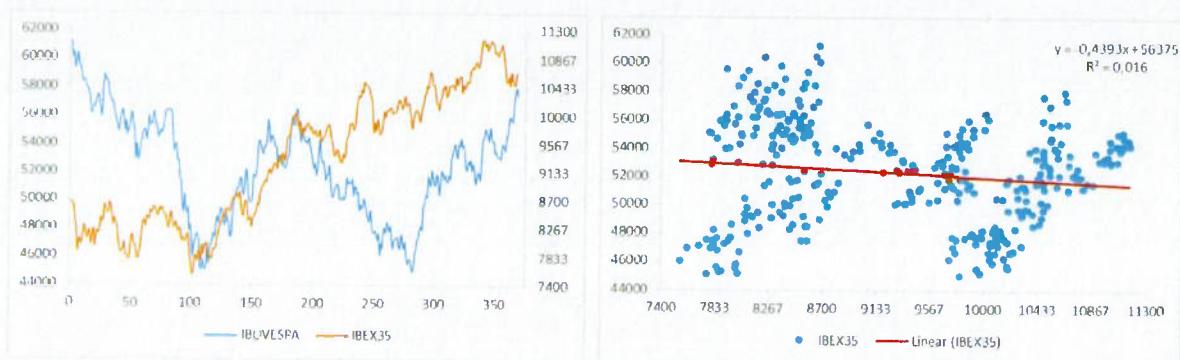


Figura 10 – Gráficos das séries temporais IBOVESPA, IBEX35 e dispersão IBOVESPA x IBEX35.

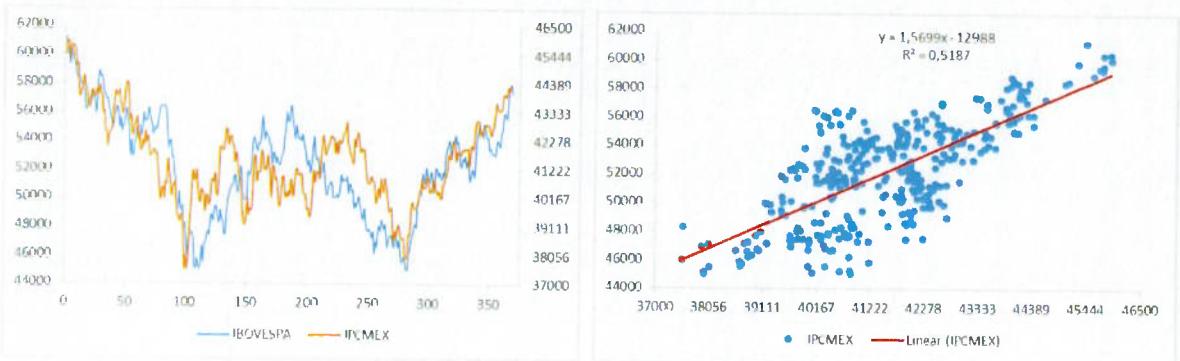


Figura 11 – Gráficos das séries temporais IBOVESPA, IPCMEX e dispersão IBOVESPA x IPCMEX.

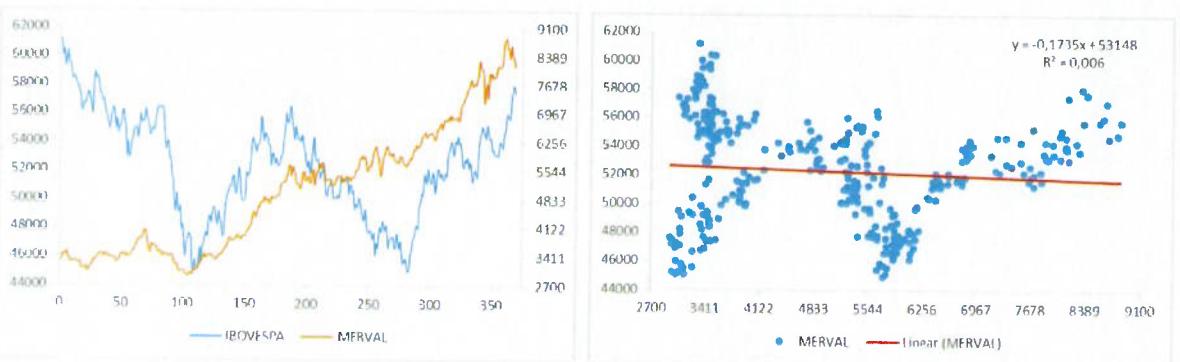


Figura 12 – Gráficos das séries temporais IBOVESPA, Merval e dispersão IBOVESPA x Merval.

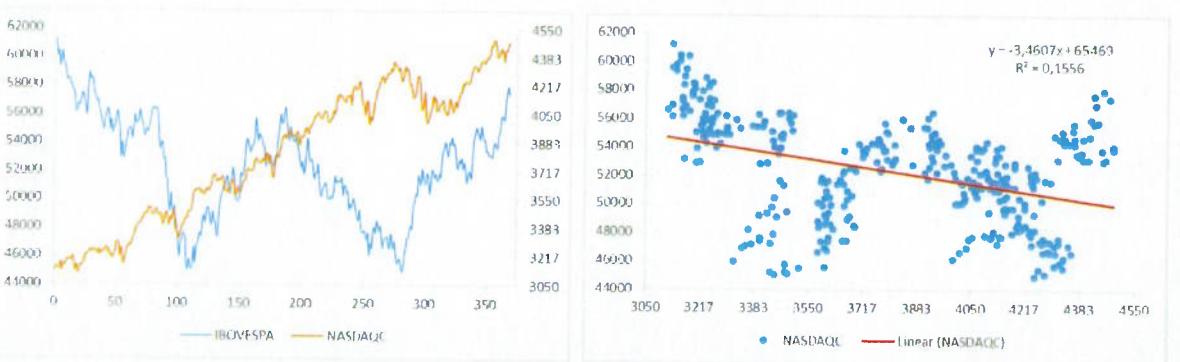


Figura 13 – Gráficos das séries temporais IBOVESPA, NASDAQC e dispersão IBOVESPA x NASDAQC.

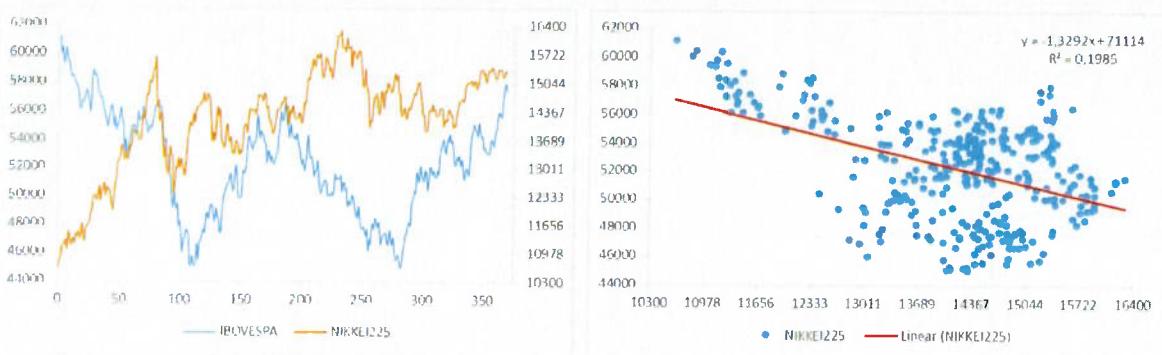


Figura 14 – Gráficos das séries temporais IBOVESPA, NIKKEI225 e dispersão IBOVESPA x NIKKEI225.

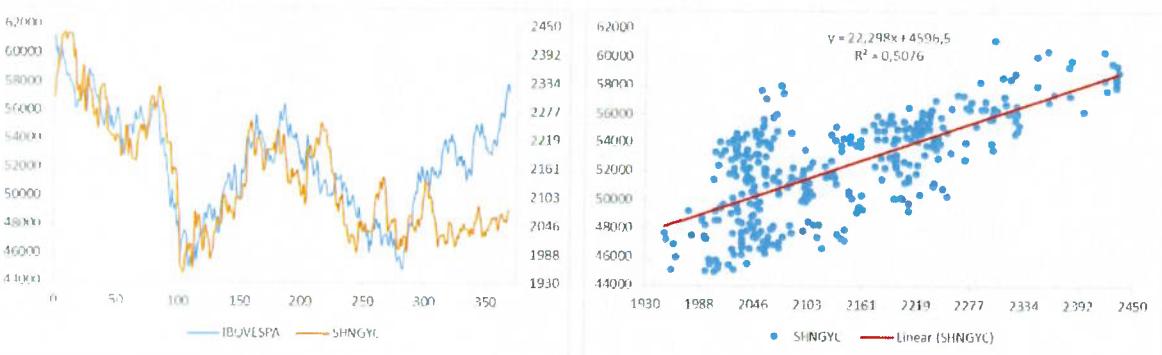


Figura 15 – Gráficos das séries temporais IBOVESPA, SHNGYC e dispersão IBOVESPA x SHNGYC.

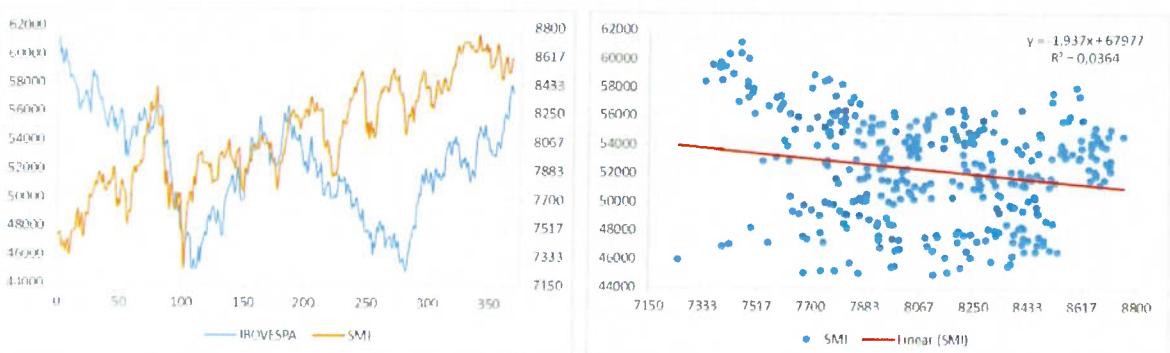


Figura 16 – Gráficos das séries temporais IBOVESPA, SMI e dispersão IBOVESPA x SMI.

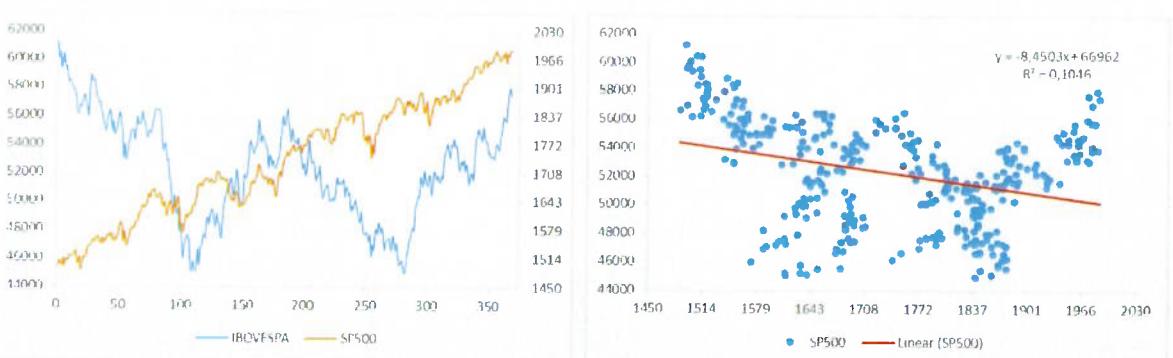


Figura 17 – Gráficos das séries temporais IBOVESPA, SP500 e dispersão IBOVESPA x SP500.

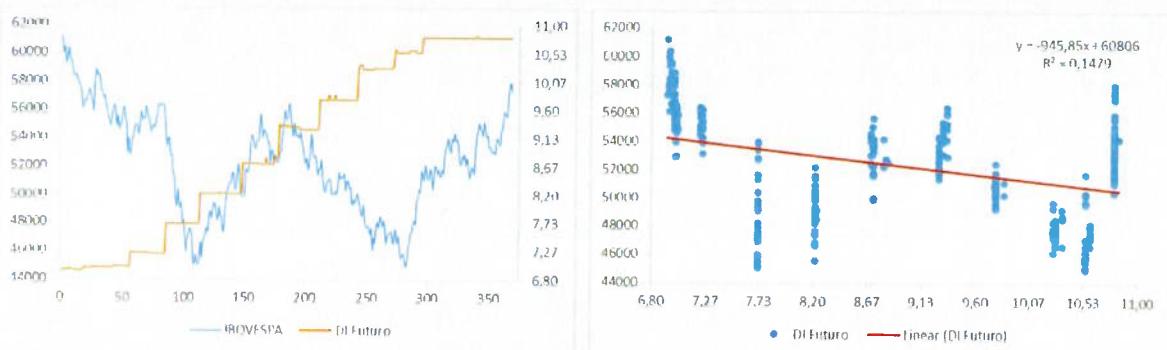


Figura 18 – Gráficos das séries temporais IBOVESPA, SELIC e dispersão IBOVESPA x SELIC.

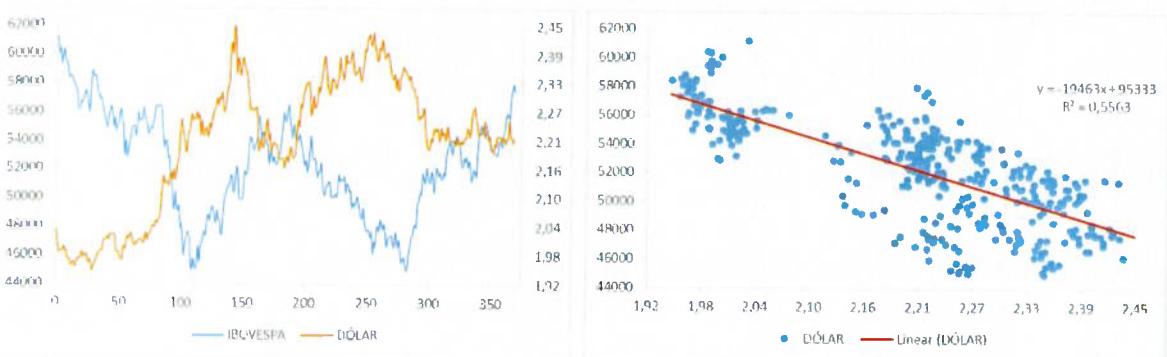


Figura 19 – Gráficos das séries temporais IBOVESPA, DÓLAR e dispersão IBOVESPA x DÓLAR.

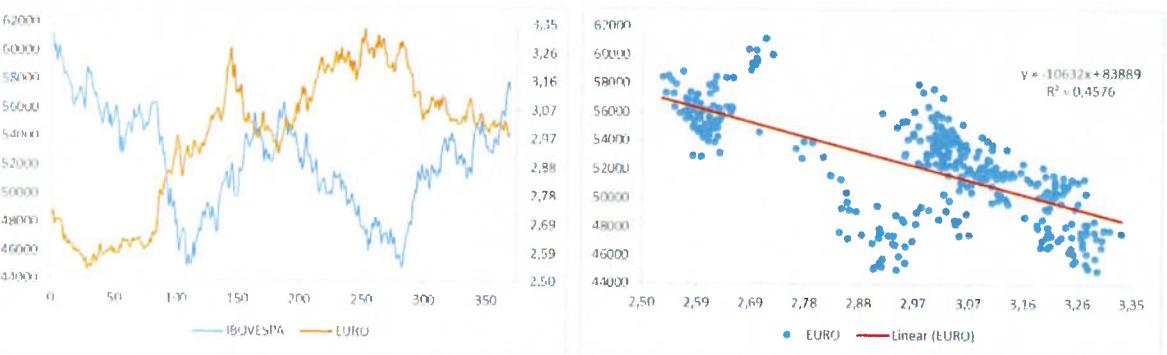


Figura 20 – Gráficos das séries temporais IBOVESPA, EURO e dispersão IBOVESPA x EURO.

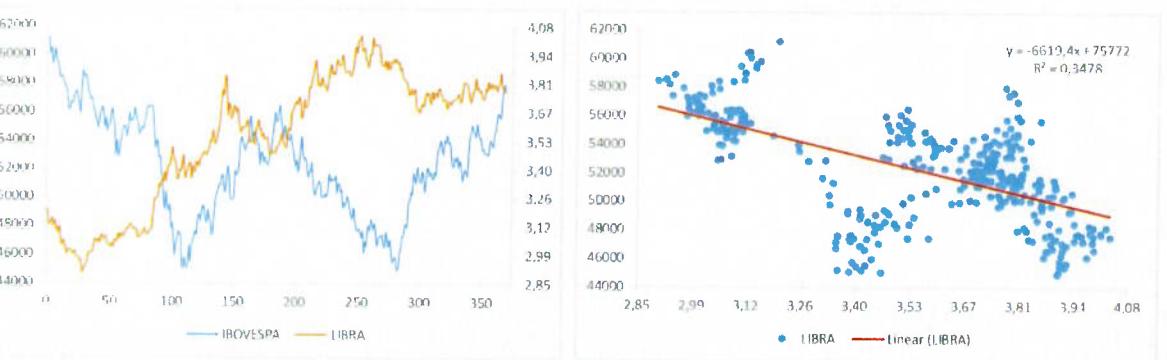


Figura 21 – Gráficos das séries temporais IBOVESPA, LIBRA e dispersão IBOVESPA x LIBRA.

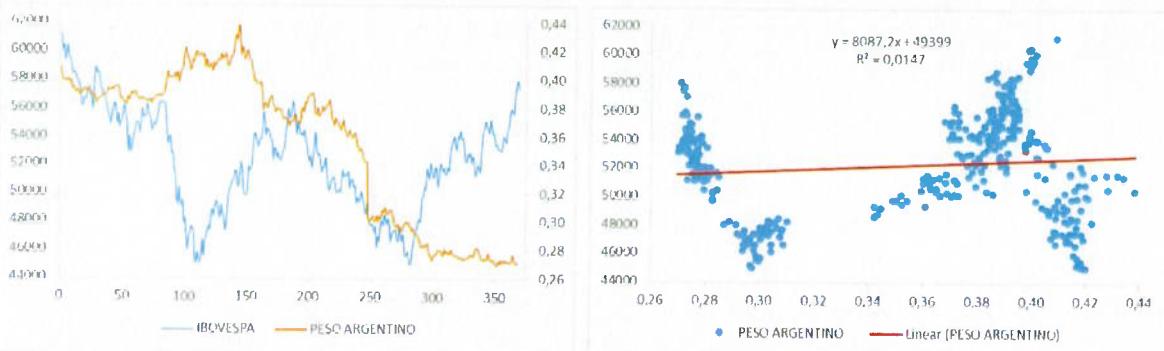


Figura 22 – Gráficos das séries temporais IBOVESPA, PESO ARGENTINO e dispersão IBOVESPA x PESO ARGENTINO.



Figura 23 – Gráficos das séries temporais IBOVESPA, IENE e dispersão IBOVESPA x IENE.

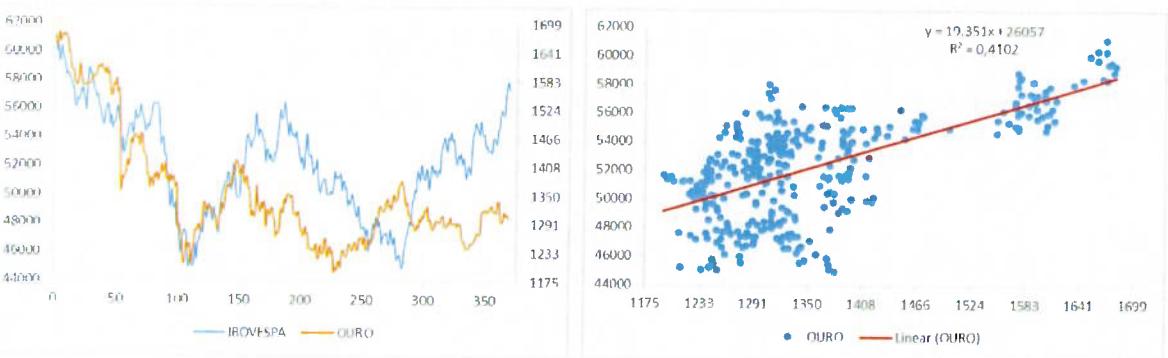


Figura 24 – Gráficos das séries temporais IBOVESPA, OURO e dispersão IBOVESPA x OURO.

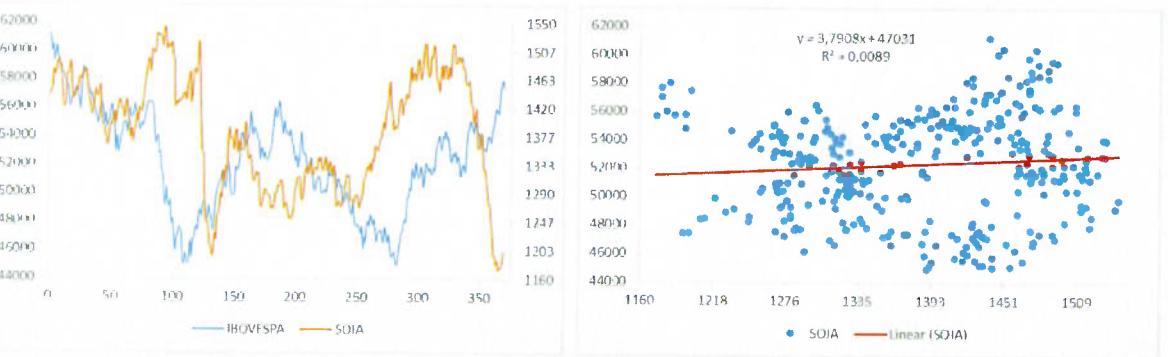


Figura 25 – Gráficos das séries temporais IBOVESPA, SOJA e dispersão IBOVESPA x SOJA.

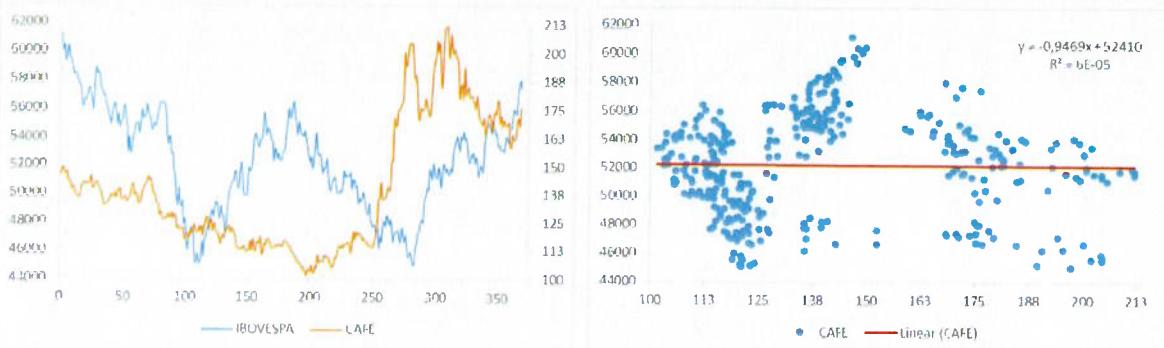


Figura 26 – Gráficos das séries temporais IBOVESPA, CAFÉ e dispersão IBOVESPA x CAFÉ.



Figura 27 – Gráficos das séries temporais IBOVESPA, PETRÓLEO e dispersão IBOVESPA x PETRÓLEO.

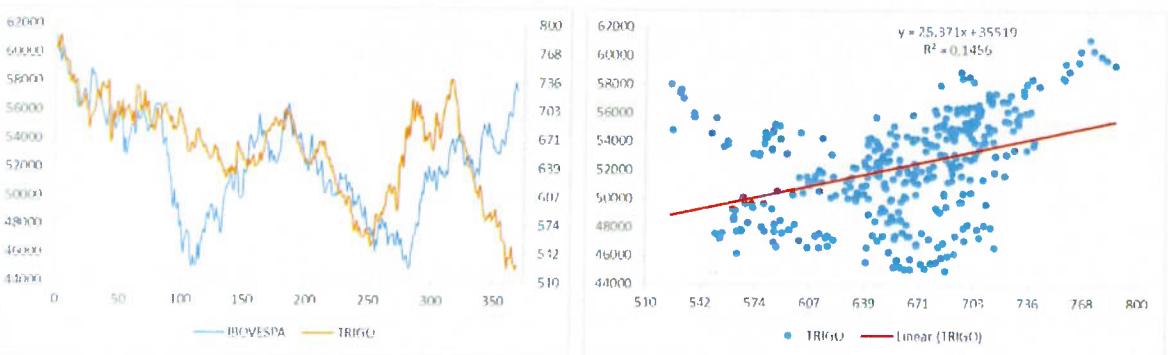


Figura 28 – Gráficos das séries temporais IBOVESPA, TRIGO e dispersão IBOVESPA x TRIGO.

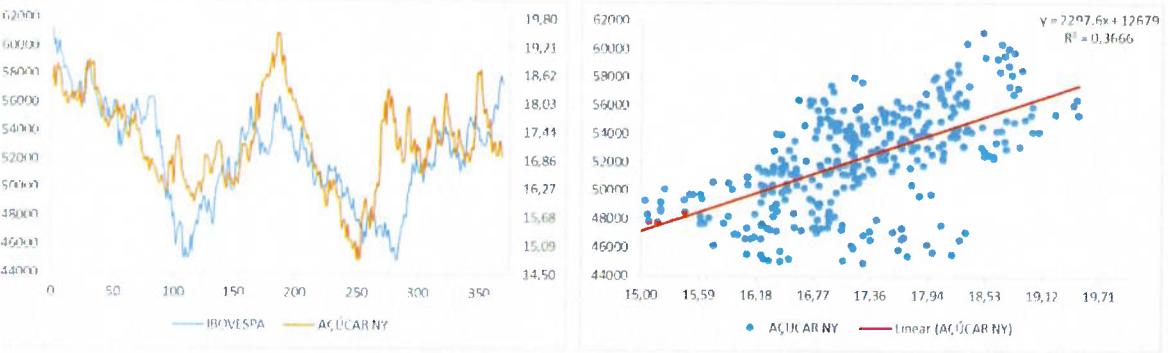


Figura 29 – Gráficos das séries temporais IBOVESPA, AÇÚCAR e dispersão IBOVESPA x AÇÚCAR.

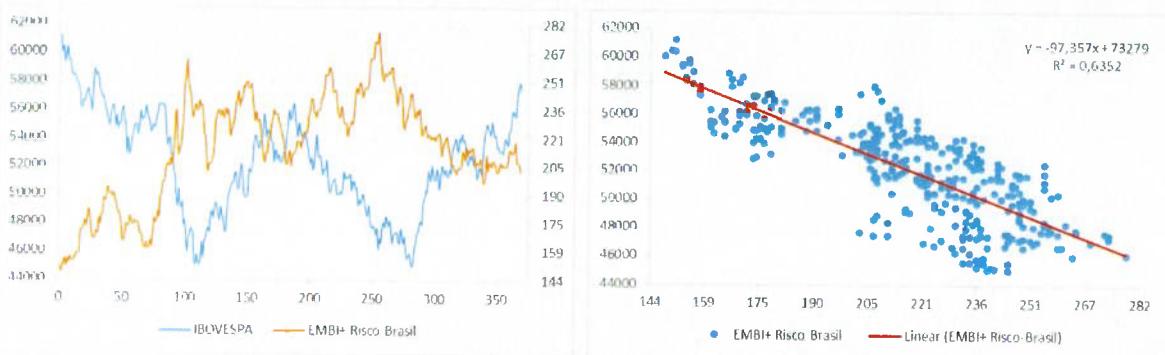


Figura 30 – Gráficos das séries temporais IBOVESPA, EMBI+Risco Brasil e dispersão IBOVESPA x EMBI+ Risco Brasil.

Após uma análise visual dos gráficos, é possível perceber que existe uma proximidade dos padrões dos índices HSENG, IPCMEX, SHNGYC, OURO e AÇÚCAR NY com o IBOVESPA. E, os índices do DÓLAR, EURO, LIBRA, IENE, PETRÓLEO, TRIGO e EMBI+Risco Brasil apresentam um padrão similar mas com um deslocamento adiantado em relação ao IBOVESPA. Portanto, os treze índices mencionados são bons candidatos ao modelo de regressão múltipla que explica o IBOVESPA.

4.2 SIMULAÇÕES DAS REGRESSÕES ATRAVÉS DO MINITAB

Através da funcionalidade *Análise de Correlação* do Excel foi gerada a Tabela 4, onde são indicadas todas as correlações das varáveis com o IBOVESPA. E através desta tabela, foram selecionados quatro subconjuntos de variáveis explanatórias para elaboração de quatro modelos de regressão múltipla. Os subconjuntos possuem vinte e seis, vinte, quinze e dez variáveis respectivamente, e através do Minitab foram verificados os melhores modelos de regressão múltipla para cada subconjunto selecionado.

Os quatro modelos calculam o valor do IBOVESPA no mesmo dia que os demais índices apresentam seu preço de fechamento. Esses modelos servem para explicar como a combinação de diversos índices influencia o movimento do IBOVESPA. Normalmente, esse tipo de modelo não é utilizado para prever Y , a menos que os valores futuros das variáveis explanatórias sejam conhecidos.

De fato, nesse trabalho, os valores futuros das variáveis explanatórias não são conhecidos, mas a fim de se obter previsões do IBOVESPA são elaborados mais dois

modelos, com base no conjunto de dez variáveis, onde o primeiro considera todas as variáveis explanatórias com um atraso de cinco dias, e o segundo, atrasos variados de acordo com o aumento das correlações entre as variáveis e o IBOVESPA.

Nesses dois modelos, pretende-se verificar se existe alguma reação futura do IBOVESPA em relação aos valores passados das variáveis explanatórias.

Tabela 4 – Correlações dos vinte e seis índices com o IBOVESPA, calculados através do Excel.

ÍNDICE	FATOR DE CORRELAÇÃO	ÍNDICE	FATOR DE CORRELAÇÃO
IPCMEX	0,7202	CAC40	-0,2240
SHNGYC	0,7124	DAX30	-0,2833
OURO	0,6404	PETRÓLEO	-0,2903
AÇÚCAR NY	0,6055	SP500	-0,3234
HSENG	0,5846	DJI	-0,3604
TRIGO	0,3815	SELIC	-0,3882
PESO ARGENTINO	0,1214	NASDAQC	-0,3944
SOJA	0,0941	NIKKEI225	-0,4456
CAFÉ	-0,0078	LIBRA	-0,5898
MERVAL	-0,0778	IENE	-0,6336
IBEX35	-0,1266	EURO	-0,6765
FTSEMIB	-0,1479	DÓLAR	-0,7459
SMI	-0,1909	EMBI+Risco Brasil	-0,7970

4.2.1 Simulação com o conjunto de vinte e seis variáveis

Uma das funcionalidades do Minitab se chama *Best Subsets*, que testa de forma automatizada combinações das variáveis explanatórias e indica o melhor modelo que se ajusta ao conjunto das observações.

Através dessa funcionalidade, foram realizados estudos das possíveis combinações das variáveis de cada subconjunto.

A Tabela 5 mostra o resultado que o Minitab gera após a execução da função *Best Subsets*. Na primeira simulação foram usadas todas as variáveis da Tabela 4.

Uma característica importante da função *Best Subsets* é que ela testa todas as combinações possíveis e apresenta, por padrão, os dois melhores resultados para cada N , onde N é o número de variáveis combinadas que foram testadas. Mas para

simplificar a tabela com os resultados, neste trabalho é apresentado apenas um modelo para cada N , o que não compromete no resultado final.

Na Tabela 5 são apresentadas trinta e duas colunas, onde as seis primeiras apresentam informações analíticas de cada modelo, e as outras vinte e seis colunas mostram quais variáveis foram selecionadas em cada análise.

Tabela 5 – Resultado da função *Best Subsets* do Minitab para o conjunto de vinte e seis variáveis.

A seguir são dadas explicações sobre as seis primeiras colunas, representadas por *Vars*, *R-Sq*, *R-Sq (adj)*, *R-Sq (pred)*, *Mallows' Cp* e *S*.

- ✓ Vars – quantidade de variáveis explanatórias que foram selecionadas para a

análise do modelo;

- ✓ *R-Sq* – representa o coeficiente de determinação amostral, e é uma medida estatística que explica o quão próximo o modelo linear se ajusta ao conjunto das observações;
- ✓ *R-Sq (adj)* – esse coeficiente é uma versão melhorada do coeficiente anterior para se ajustar à quantidade de variáveis do modelo e evitar modelos tendenciosos. A desvantagem do coeficiente anterior em relação a este é que ele poderia aumentar com o acréscimo de novas variáveis, enquanto o *R-Sq (adj)* aumenta apenas se a variável adicional for estatisticamente significante para o modelo. Na Tabela 6 é apresentado um exemplo de como essa variação pode ocorrer;

Tabela 6 – Comparação dos coeficientes *R-Sq* e *R-Sq (adj)* da função *Best Subsets* do Minitab.

Vars	R-Sq	R-Sq (adj)
1,00	72,10	71,00
2,00	85,90	84,80
3,00	87,40	85,90
4,00	89,10	82,30
5,00	89,90	80,70

- ✓ *R-Sq (pred)* – esse coeficiente indica a qualidade das respostas que o modelo de regressão gera para novas observações, e o benefício deste é que ele auxilia na identificação de modelos sobre estimados. Se esse coeficiente for muito menor que os anteriores, significa que o modelo possui mais variáveis explanatórias que o necessário.
- ✓ *Mallows' Cp* – é um indicador que compara a precisão e tendências entre modelos utilizando combinações distintas das variáveis explanatórias. De acordo com o site de suporte do Minitab, deve-se buscar por modelos com valores pequenos e próximos à quantidade de variáveis independentes mais um. Modelos que respeitam esse critério, apresentam uma precisão maior com relação ao dado previsto e observado;
- ✓ *S* – é uma medida de erro padrão, ou seja, ele mostra qual é a distância média

entre os valores observados e os valores previstos. Valores menores indicam um modelo mais aderente ao conjunto das observações.

De acordo com os resultados da Tabela 5, o melhor modelo é aquele que contém dezoito variáveis. O mesmo apresenta o melhor *R-Sq (adj)*, o menor *Mallows' Cp* e o menor *S* respectivamente. As variáveis selecionadas para este modelo estão assinaladas com um "X" nas suas respectivas colunas.

Desenvolvendo o modelo de regressão com dezoito variáveis através do Minitab, chega-se ao resultado da Tabela 7. Para simplificar as explicações, o modelo de regressão gerado é citado como modelo (18) daqui em diante.

Tabela 7 – Modelo de regressão múltipla com 18 variáveis gerado pelo Minitab.

Term	Coef	SE Coef	T-Value	P-Value	VIF
CAC40	10,010	1,22	8,21	0,000	45,46
DAX30	-1,542	0,58	-2,67	0,008	79,13
HSENG	0,446	0,19	2,32	0,021	9,61
IPCMEX	0,117	0,07	1,73	0,084	5,04
MERVAL	1,654	0,18	9,05	0,000	35,32
NASDAQC	-10,770	1,41	-7,62	0,000	136,21
NIKKEI225	-0,678	0,20	-3,43	0,001	23,02
SHNGYC	7,180	1,40	5,12	0,000	10,56
SMI	-1,116	0,69	-1,63	0,105	24,06
SP500	17,210	4,17	4,12	0,000	134,01
LIBRA	3293,000	1211,00	2,72	0,007	61,23
PESO ARGENTINO	15572,000	6108,00	2,55	0,011	44,16
IENE	-1247044,000	241896,00	-5,16	0,000	27,19
OURO	9,650	1,41	6,86	0,000	11,38
SOJA	-3,130	1,16	-2,69	0,007	4,38
CAFE	-37,250	6,28	-5,93	0,000	14,01
TRIGO	-5,250	2,25	-2,33	0,020	6,02
EMBI+Risco-Brasil	-23,920	6,33	-3,78	0,000	14,13
Constant	31302,000	7193,00	4,35	0,000	

Na Tabela 7 estão as dezoito variáveis selecionadas mais o valor da constante do modelo. A seguir são dadas as explicações sobre as colunas dessa tabela, representadas por *Term*, *Coef*, *SE Coef*, *T-Value*, *P-Value* e *VIF*.

- ✓ *Term* – identifica as variáveis selecionadas para o modelo. A última linha se refere à constante do modelo;
- ✓ *Coef* – apresenta os valores calculados para os coeficientes de cada variável explanatória, assim como o valor da constante;
- ✓ *SE Coef* – apresenta o erro padrão do coeficiente estimado. Quanto menor o valor do erro padrão, mais precisa é a estimativa do coeficiente;
- ✓ *T-Value* – indica o nível de significância do coeficiente para o modelo. Quanto mais distante de zero, mais significante é a variável para o modelo;
- ✓ *P-Value* – é utilizado para validar a hipótese nula de que o coeficiente é igual a zero (nenhum efeito). Um *p-value* baixo ($< 0,05$) indica que se pode rejeitar a hipótese nula e que a adição da variável é significativa para o modelo.
- ✓ *VIF* – usado para descrever quanta multicolinearidade a variável apresenta em relação às outras variáveis do modelo. A multicolinearidade é problemática porque pode aumentar o erro padrão dos coeficientes de regressão, tornando-os instáveis e difíceis de interpretar. A Tabela 8 pode ser utilizada para interpretar os valores de VIF.

Tabela 8 – Referência de valores para interpretação do VIF (FROST, 2014).

VIF	Status das variáveis explanatórias
VIF = 1	Não correlacionados
$1 < VIF < 5$	Moderadamente correlacionados
$VIF > 5$	Altamente correlacionados

Com base nos valores da Tabela 7, a seguir é representado a equação matemática do modelo (18).

$$\begin{aligned}
 IBOVESPA_t = & 31302 + 10,01 CAC40_t - 1,542 DAX30_t + 0,446 HSENG_t + 0,1168 IPCMEX_t + \\
 & 1,654 Merval_t - 10,77 NASDAQC_t - 0,678 NIKKEI225_t + 7,18 SHNGYC_t - 1,116 SMI_t + 17,21 SP500_t + \\
 & 3293 LIBRA_t + 15572 PESO ARGENTINO_t - 1247044 IENE_t + 9,65 OURO_t - 3,13 SOJA_t - 37,25 CAFE_t - \\
 & 5,25 TRIGO_t - 23,92 EMBI RISCO BRASIL_t \quad (4-1)
 \end{aligned}$$

A partir do conjunto de observações e do modelo (18), foi plotado o gráfico da

Figura 31, onde a curva em azul representa a série histórica do IBOVESPA no intervalo entre os dias 23/01/2013 e 23/01/2015, a curva em laranja os valores ajustados através do modelo (18), no intervalo entre os dias 23/01/2013 e 23/07/2014² e a curva em vermelho, os valores previstos³ no intervalo entre os dias 24/07/2014 e 23/01/2015, período utilizado para validar o modelo.



Figura 31 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (18).

É possível observar que os dados ajustados pelo modelo (18) estão bem aderentes aos dados observados no período utilizado para a elaboração do modelo, mas rapidamente, quando se inicia o período para validação, percebe-se que o mesmo perde sua capacidade de estimar o valor do IBOVESPA.

4.2.1.1 Análise dos resíduos e modelo ARMA (p, q)

A partir da diferença entre os valores observados e os valores ajustados do modelo (18), pode-se obter os resíduos. Na Tabela 9, seguem os valores iniciais da série temporal do IBOVESPA, os valores ajustados e os resíduos, correspondentes ao mesmo período.

Os trabalhos de (NETO, 2006) e (MEDEIROS, 2006) explicam como os resíduos gerados a partir de uma regressão múltipla são utilizados para explicar o nível

² O intervalo entre os dias 23/01/2013 e 23/07/2014 foi utilizado para elaboração de todos os modelos.

³ Os valores previstos não são valores futuros do IBOVESPA, são respostas do modelo para o período de testes, período que não foi utilizado para elaboração do modelo.

de confiança dos resultados do modelo. De acordo com os autores, se o modelo for apropriado, a série temporal dos resíduos apresenta propriedades que a aproxima de uma distribuição normal, num caso ideal, a série dos resíduos poderia ser considerada como um ruído branco.

Tabela 9 – Valores iniciais da série histórica do IBOVESPA, os valores ajustados pelo modelo (18) e os resíduos calculados.

Data	IBOVESPA	Valores	
		Ajustados Modelo (18)	Resíduos
24/01/2013	61169,000	58599,447	2569,553
28/01/2013	60027,000	58892,706	1134,294
29/01/2013	60406,000	59447,413	958,587
30/01/2013	59336,000	59673,719	-337,719
31/01/2013	59761,000	59112,218	648,782
01/02/2013	60351,000	60215,261	135,739
04/02/2013	59575,000	59410,656	164,344
05/02/2013	59444,000	59691,024	-247,024
06/02/2013	58951,000	58967,286	-16,286
07/02/2013	58372,000	58639,619	-267,619

No seu trabalho, (NETO, 2006) faz uma análise da série temporal dos resíduos e chega a um modelo ARMA, que é somado ao modelo inicial. Neste trabalho é adotada a mesma abordagem para obtenção de melhores resultados.

Seguindo este conceito, o Minitab foi configurado para apresentar, como saída dos estudos, uma análise dos resíduos do modelo de regressão. Na Figura 32 é apresentado o comportamento da série temporal dos resíduos do modelo (18) ao longo do tempo, indicando como eles oscilam ao redor do zero, sem apresentar nenhuma tendência definida, o que por si só, mostra que a autocorrelação entre os resíduos é fraca. De acordo com (GUJARATI e PORTER, 2011), séries temporais consideradas estacionárias apresentam distribuição normal $N(0, \sigma^2)$.

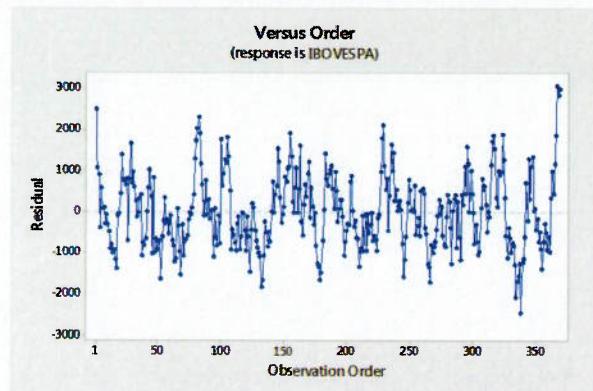


Figura 32 – Resíduos do modelo (18) em relação ao tempo.

Na Figura 33 é possível verificar que o histograma dos resíduos também se aproxima bastante de um histograma com distribuição normal, validando a observação anterior.

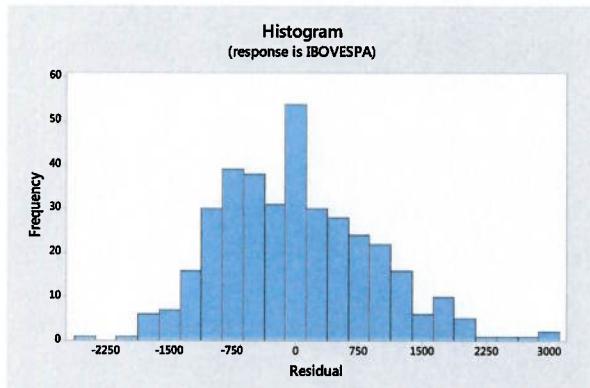


Figura 33 – Histograma dos resíduos do modelo (18).

O gráfico de resíduos versus os valores ajustados, além de indicar se há heteroscedasticidade, pode indicar que não existe uma relação linear entre as variáveis explanatórias com a variável dependente por meio de alguma tendência nos pontos plotados. Na Figura 34 é possível observar que os pontos estão aleatoriamente distribuídos em torno do zero, sem nenhum comportamento ou tendência definidos. A partir dessa observação é possível dizer que a variância dos resíduos é homoscedástica, o que garante que a inferência estatística é válida e que a estimativa dos coeficientes do modelo, através do método Mínimos Quadrados Ordinários, foram corretas.

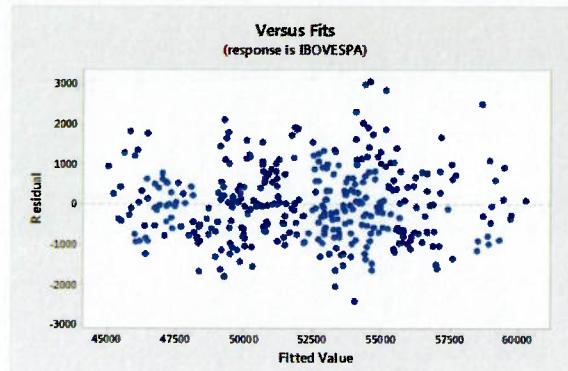


Figura 34 – Resíduos versus valores ajustados do modelo (18).

A Figura 35 mostra o gráfico normal de probabilidade dos resíduos muito próximo à reta de referência, o que também é uma característica de conjuntos de dados que apresentam distribuição normal.

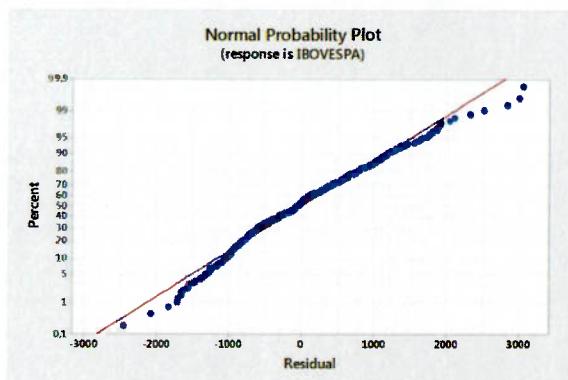


Figura 35 – Gráfico normal de probabilidade, ou *quantil quantil*, dos resíduos do modelo (18).

Após a análise dos resíduos, pode-se dizer que a série temporal destes é estacionária e possui distribuição normal. E como os dados ajustados estão aderentes aos dados observados, ficam validadas as observações feitas por (NETO, 2006) e (MEDEIROS, 2006) quando dizem que a qualidade dos resíduos indicam o nível de confiança do modelo de regressão.

Na Figura 36 é possível verificar um nível de aderência de 93,34% no período de elaboração do modelo, o que vai de encontro ao mencionado no parágrafo anterior. Mas para o período de testes as observações anteriores não são válidas, uma vez que o nível de aderência está em torno de 5%.

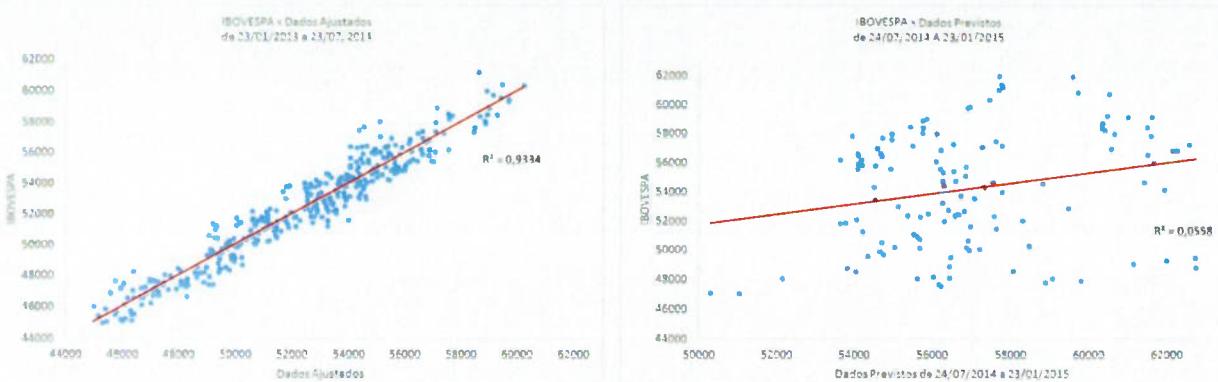


Figura 36 – Gráficos de dispersão para comparar a qualidade das respostas e os valores observados entre os períodos de ajuste e teste do modelo (18).

Como parte da análise deste trabalho, e conforme realizado no trabalho de (NETO, 2006), é elaborado um modelo ARMA que descreve os resíduos e este é somado ao modelo principal para aumentar a qualidade das respostas.

De acordo com os gráficos de autocorrelação e autocorrelação parcial dos resíduos do modelo (18), representados pelas Figuras 37 e 38, chegou-se à conclusão⁴ de que o melhor modelo para descrever o movimento dos resíduos é um AR(1), representado na Tabela 10.

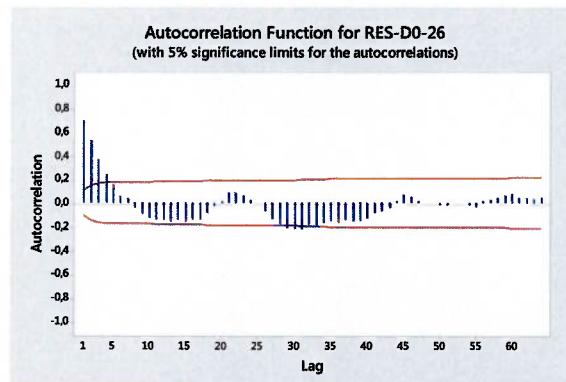


Figura 37 – Gráfico da função de autocorrelação dos resíduos do modelo (18).

⁴ Os gráficos de autocorrelação e autocorrelação parcial são ferramentas fundamentais para identificar o modelo ARMA que melhor descreve a série temporal em questão. Não foram abordados aspectos matemáticos sobre o tema, uma vez que esse procedimento é comum em literaturas sobre modelos de séries temporais.

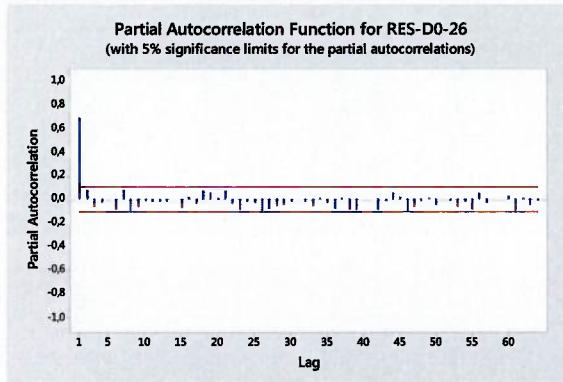


Figura 38 – Gráfico da função de autocorrelação parcial dos resíduos do modelo (18).

Utilizando a funcionalidade de análise de séries temporais do Minitab, foi elaborado um modelo AR de ordem *um* para o conjunto de resíduos do modelo (18), representado pela Tabela 10.

As colunas desta tabela indicam o tipo e a ordem do coeficiente, o valor do coeficiente, o erro padrão do coeficiente e as estatísticas que indicam o nível de significância do coeficiente em questão.

Tabela 10 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (18).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,7380	0,0364	20,26	0,0000
Constant	24,8500	32,8400	0,76	0,4500

Analizando os dados da Tabela 10, percebe-se que a constante não é uma variável significativa para o modelo, portanto não será considerada como parte do modelo AR(1), representado pela equação matemática a seguir.

$$Res_t = 0,7380 Res_{t-1} \quad (4-2)$$

Uma vez elaborado o modelo (4-2), o mesmo foi adicionado ao modelo (18) o que gerou uma melhora considerável nas respostas para o período de testes, e que pode ser evidenciado nos gráficos das Figuras 39 e 40.

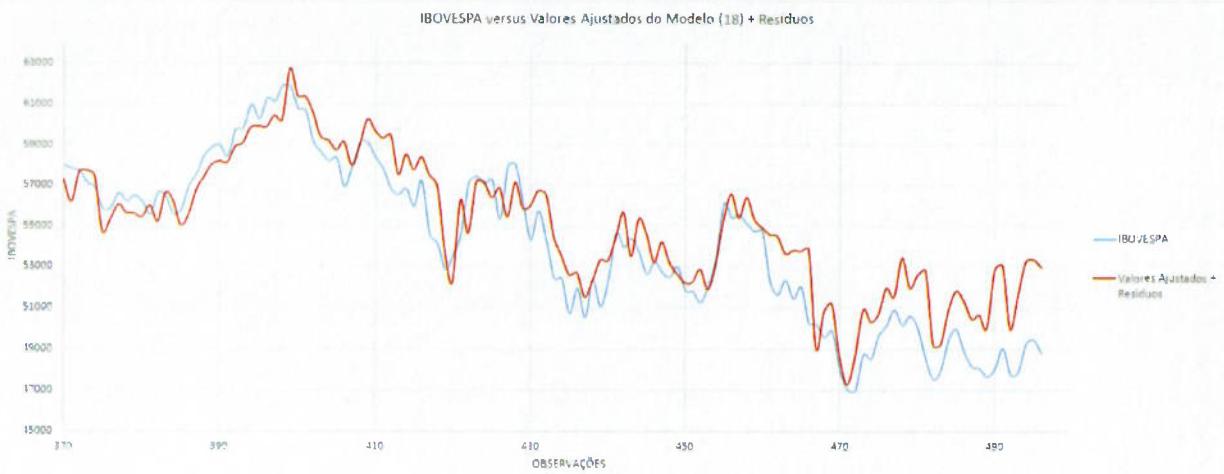


Figura 39 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (18) com a adição do modelo auto regressivo dos resíduos (4-2).

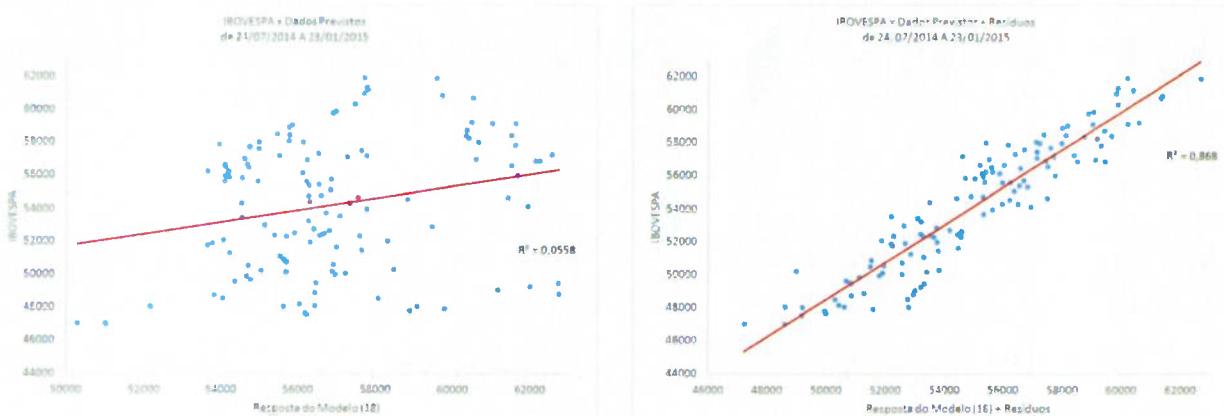


Figura 40 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (18) antes e depois da adição do modelo auto regressivo dos resíduos (4-2).

4.2.2 Simulação com o conjunto de vinte variáveis

Seguindo com a elaboração dos próximos modelos, com vinte, quinze e dez variáveis explanatórias, os resultados são apresentados de forma mais simplificada para que este trabalho não fique repetitivo.

A seguir, a função *Best Subsets* foi executada com um número menor de variáveis, onde foram desconsideradas as seis menos relevantes, ou seja, aquelas que apresentavam uma correlação próxima a zero, totalizando vinte variáveis. A Tabela 11 mostra os índices utilizados para gerar as combinações do *Best Subset*, que são apresentadas na Tabela 12.

Tabela 11 - Correlações dos vinte índices mais relevantes com o IBOVESPA.

ÍNDICE	FATOR DE CORRELAÇÃO	ÍNDICE	FATOR DE CORRELAÇÃO
IPCMEX	0,7202	SP500	-0,3234
SHNGYC	0,7124	DJI	-0,3604
OURO	0,6404	SELIC	-0,3882
AÇÚCAR NY	0,6055	NASDAQC	-0,3944
HSENG	0,5846	NIKKEI225	-0,4456
TRIGO	0,3815	LIBRA	-0,5898
SMI	-0,1909	IENE	-0,6336
CAC40	-0,2240	EURO	-0,6765
DAX30	-0,2833	DÓLAR	-0,7459
PETRÓLEO	-0,2903	EMBI+Risco Brasil	-0,7970

Tabela 12 – Resultado da função *Best Subsets* do Minitab para o conjunto de vinte variáveis.

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows	R																							
					Cp	S	X	C	O	Y	R	N	N	G	M	4	3	E	0	J	I	Q	2	R	N	R	A	I
1	63,5	63,40	63,20	958,60	2132,70																							X
2	81,7	81,60	81,40	300,70	1512,30						X																	X
3	83,3	83,10	82,90	245,80	1448,00						X																	X
4	84,5	84,40	84,10	202,20	1394,40						X					X											X	
5	85,9	85,70	85,30	155,40	1334,30	X					X					X											X	
6	86,7	86,50	86,10	128,00	1297,20	X					X					X											X	
7	87,4	87,20	86,80	102,90	1262,00	X					X					X	X			X							X	
8	88,0	87,70	87,30	84,30	1235,00	X	X				X					X		X	X	X							X	
9	88,4	88,10	87,70	72,50	1216,80	X	X				X	X				X		X	X	X							X	
10	88,8	88,50	88,10	58,10	1194,60	X	X				X	X	X			X		X	X	X							X	
11	89,1	88,80	88,30	49,50	1180,60	X	X				X	X				X	X	X	X	X							X	
12	89,4	89,00	88,60	41,70	1167,40	X	X				X	X	X	X		X	X	X	X	X							X	
13	89,8	89,40	88,90	29,40	1147,20	X	X	X			X	X	X	X		X	X	X	X	X							X	
14	90,0	89,60	89,10	23,40	1136,30	X	X	X			X	X	X	X		X	X	X	X	X		X	X			X		
15	90,1	89,70	89,10	21,60	1132,00	X	X	X			X	X	X	X		X	X	X	X	X		X	X	X		X		
16	90,2	89,80	89,20	19,40	1126,80	X	X	X			X	X	X	X		X	X	X	X	X		X	X	X		X		
17	90,3	89,90	89,30	17,70	1122,60	X	X	X			X	X	X	X		X	X	X	X	X		X	X	X		X		
18	90,4	89,90	89,20	18,30	1121,80	X	X	X			X	X	X	X		X	X	X	X	X		X	X	X		X		
19	90,4	89,90	89,20	19,30	1122,00	X	X	X			X	X	X	X		X	X	X	X	X		X	X	X		X		
20	90,4	89,90	89,10	21,00	1123,00	X	X	X	X		X	X	X	X		X	X	X	X	X		X	X	X		X		

De acordo com os resultados da Tabela 12, o melhor modelo é o que contém dezessete variáveis, denominado modelo (17) daqui em diante, e é possível verificar que nesse modelo o valor de *R-Sq (adj)* reduziu 3,97% em relação ao modelo (18), que apresenta um *R-Sq (adj)* de 93,00%. Os outros parâmetros também apresentaram uma piora se comparados ao modelo anterior.

A Tabela 13 apresenta os coeficientes, calculados pelo Minitab, do modelo (17) e é possível perceber que foram mantidas doze variáveis do modelo anterior, e sete dessas variáveis apresentam uma redução do *VIF* (multicolinearidade).

Das cinco variáveis que foram adicionadas ao último modelo, quatro apresentam valores elevados do *VIF*, o que justifica o aumento de algumas variáveis que se repetem nos dois modelos, como por exemplo a SP500, que teve um aumento do *VIF* em 300%.

Tabela 13 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo (17) variáveis.

Term	Coef	SE Coef	T-Value	P-Value	VIF
CAC40	16,230	1,50	3,28	0,001	0,00
DAX30	-2,283	0,63	-3,60	0,000	66,08
DJI	-4,474	0,80	-5,57	0,000	134,95
HSENG	0,769	0,21	3,68	0,000	7,79
IPCMEX	0,428	0,07	5,95	0,000	3,94
NASDAQC	-15,450	1,84	-8,41	0,000	158,99
NIKKEI225	0,475	0,18	2,61	0,009	13,47
SHNGYC	7,570	1,56	4,86	0,000	8,99
SMI	-4,142	0,88	-4,69	0,000	27,47
SP500	65,160	8,76	7,44	0,000	408,01
DOLAR	-8950,000	3805,00	-2,35	0,019	77,22
EURO	-5145,000	2686,00	-1,92	0,056	106,06
SELIC	-700,000	307,00	-2,28	0,023	54,12
LIBRA	6020,000	1959,00	3,07	0,002	110,62
OURO	6,190	1,59	3,90	0,000	10,04
PETROLEO	-77,700	17,90	-4,35	0,000	2,33
TRIGO	-14,810	2,78	-5,32	0,000	6,36
Constant	25800,000	7863,00	3,28	0,001	0,00

Outra observação interessante é que três variáveis que se repetem nos dois modelos, NASDAQ, SMI e SP500 são índices americanos e duas variáveis que foram

adicionadas, DJI e DÓLAR, também são índices americanos, logo é esperado que ocorra o aumento do *VIF* para essas cinco variáveis.

Com base nos coeficientes da Tabela 13, foi elaborada a equação a seguir, que representa o modelo (17).

$$\begin{aligned} IBOVESPA_t = 25800 + 16,23 CAC40_t - 2,283 DAX30_t - 4,474 DJI_t + 0,769 HSENG_t + 0,428 IPCMEX_t - \\ 15,45 NASDAQC_t + 0,475 NIKKEI225_t + 7,57 SHNGYC_t - 4,142 SMI_t + 65,16 SP500_t - 8950 DÓLAR_t - \\ 5145 EURO_t - 700 SELIC_t + 6020 LIBRA_t + 6,19 OURO_t - 77,7 PETRÓLEO_t - 14,81 TRIGO_t \quad (4-3) \end{aligned}$$

Da mesma forma que para o modelo (18), foi plotado o gráfico do modelo (17), representado na Figura 41, onde é possível verificar a aderência do modelo aos dados observados, bem como sua capacidade de previsão. Diferente do modelo anterior, a partir da observação 480 observa-se um grande afastamento das previsões em relação aos dados observados.

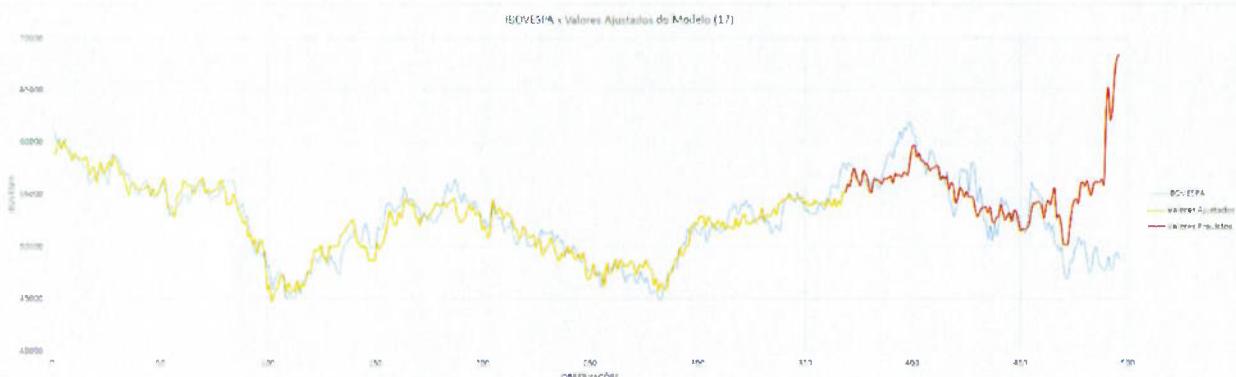


Figura 41 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (17).

4.2.2.1 Análise dos resíduos e modelo ARMA (p, q)

O Minitab também gerou os gráficos de resíduos para o modelo (17), onde é possível verificar o mesmo comportamento dos resíduos do modelo (18).

Nas Figuras 42 e 43, verifica-se, respectivamente, a evolução dos resíduos, o histograma dos resíduos, o gráfico dos resíduos versus os dados ajustados e o gráfico normal de probabilidade dos resíduos.

Assim como para o modelo anterior, é possível elaborar um modelo ARMA que descreva o movimento dos resíduos do modelo (17) e adicioná-lo ao modelo de regressão múltipla para melhorar as suas respostas.

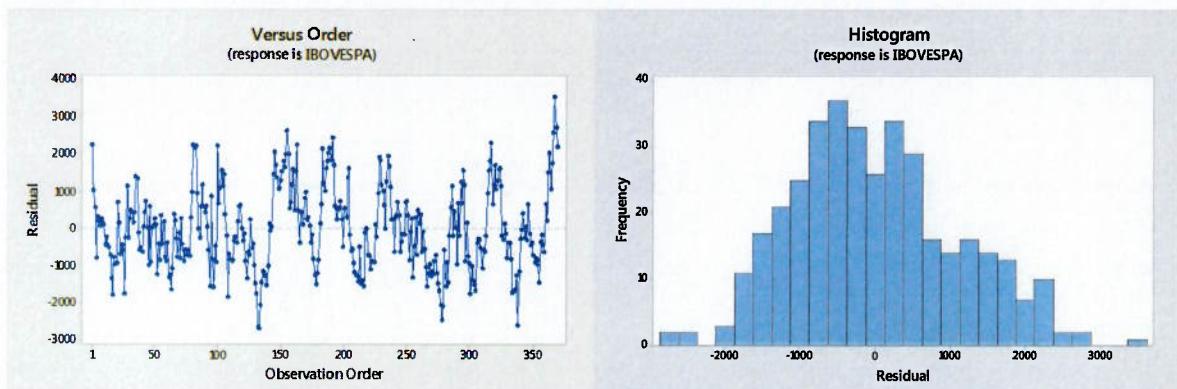


Figura 42 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (17).

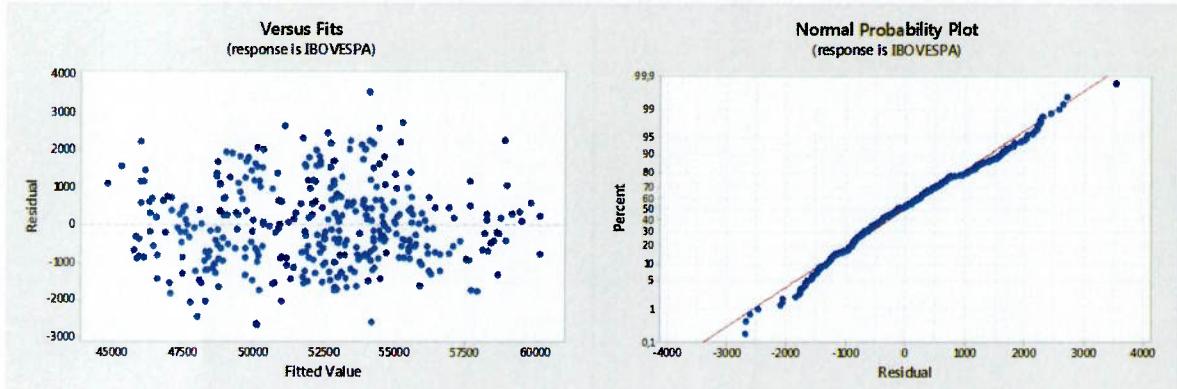


Figura 43 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.

De acordo com os gráficos de autocorrelação e o de autocorrelação parcial dos resíduos do modelo (17), representados na Figura 44, chegou-se à conclusão de que o melhor modelo para descrever o movimento dos resíduos é um AR(1). Essa conclusão se deve ao fato do gráfico da função de autocorrelação apresentar um decaimento exponencial com padrão de onda senoidal e o gráfico da função de autocorrelação parcial apresentar um pico significativo na primeira defasagem.

Utilizando a funcionalidade de análise de séries temporais do Minitab, foi elaborado o modelo AR de ordem *um* para o conjunto de resíduos do modelo (17),

representado pela Tabela 14.

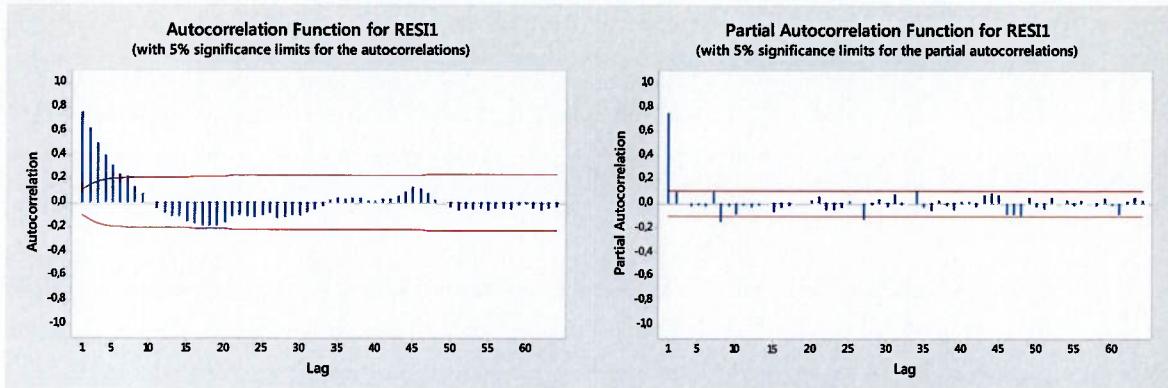


Figura 44 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (17).

Da mesma forma que ocorreu para o modelo AR anterior, a constante não é significativa e pode ser desconsiderada.

Tabela 14 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (17).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,7778	0,0333	23,34	0,0000
Constant	9,2200	36,4400	0,25	0,8010

Portanto, o modelo AR(1) para os resíduos do modelo (17), pode ser representado matematicamente conforme a equação a seguir.

$$Res_t = 0,7778 Res_{t-1} \quad (4-4)$$

A Figura 45 apresenta o gráfico do modelo (17) após a soma do modelo AR(1) (4-4) na equação (4-3).

Os modelos (17) e (18) apresentam comportamentos muito próximos um do outro, mas nos gráficos das Figuras 39 e 45 não fica claro qual deles apresenta melhor aderência aos dados observados, mas quando as Figuras 40 e 46 são comparadas, percebe-se que o modelo (18) apresenta um R^2 de 86,80% contra 83,83% do modelo (17), no período de 24/07/2014 a 23/01/2015.

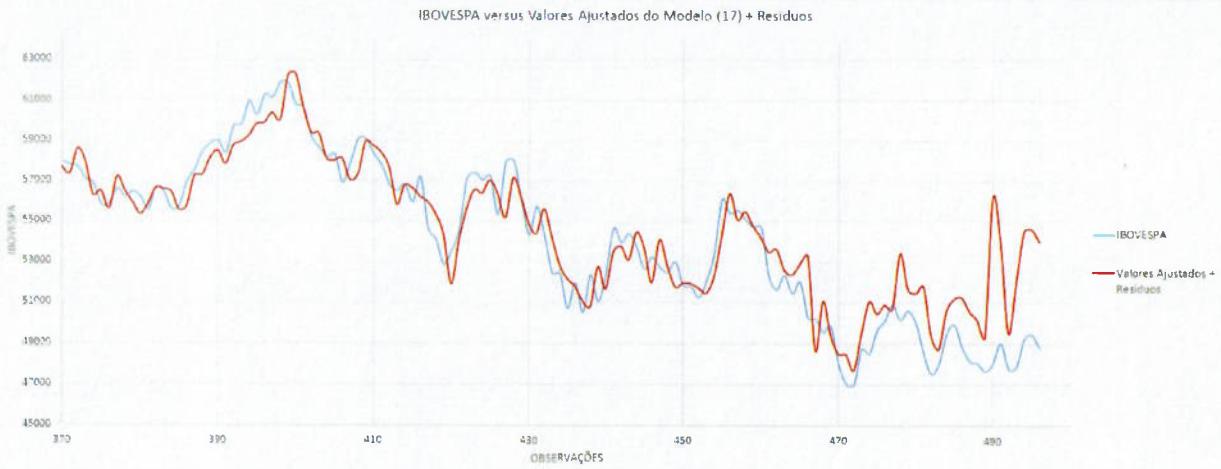


Figura 45 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (17) com a adição do modelo auto regressivo dos resíduos (4-4).

Na Figura 46 é possível perceber que o modelo (17) apresenta uma piora de 2,97% em relação ao modelo (18). Esse resultado é aceitável uma vez que houve uma redução de 3% no valor de R^2 (adj) para o modelo (17) em relação ao anterior.

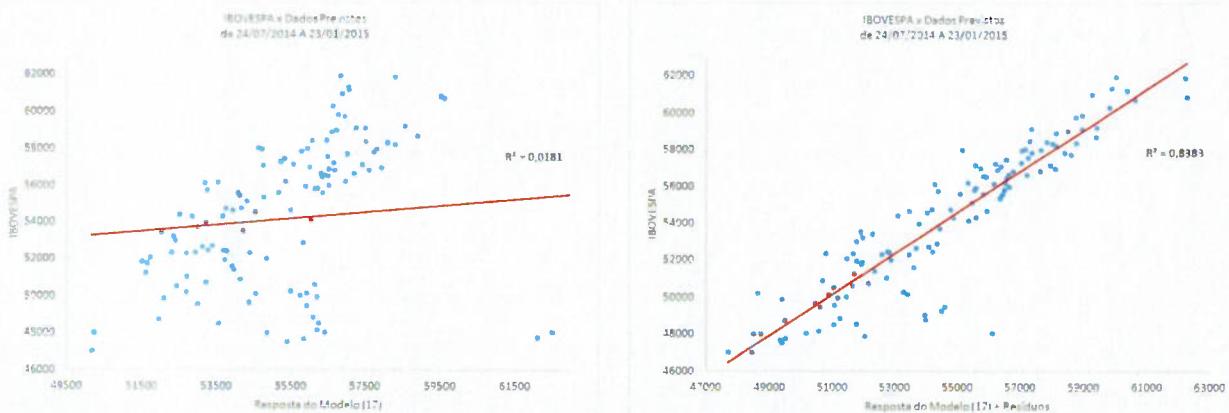


Figura 46 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (17) antes e depois da adição do modelo auto regressivo dos resíduos (4-4).

4.2.3 Simulação com o conjunto de quinze variáveis

Na terceira simulação foram desconsideradas mais cinco variáveis menos relevantes em termos de correlação com o IBOVESPA e chegou-se a um conjunto de quinze variáveis, apresentadas na Tabela 15.

Tabela 15 – Correlações dos quinze índices mais relevantes com o IBOVESPA.

ÍNDICE	FATOR DE CORRELAÇÃO	ÍNDICE	FATOR DE CORRELAÇÃO
IPCMEX	0,7202	NASDAQC	-0,3944
SHNGYC	0,7124	NIKKEI225	-0,4456
OURO	0,6404	LIBRA	-0,5898
AÇÚCAR NY	0,6055	IENE	-0,6336
HSENG	0,5846	EURO	-0,6765
TRIGO	0,3815	DÓLAR	-0,7459
DJI	-0,3604	EMBI+Risco Brasil	-0,7970
SELIC	-0,3882		

Na Tabela 16 é apresentado o resultado do *Best Subsets* do Minitab, indicando que o melhor modelo contém doze variáveis, denominado modelo (12) daqui em diante, e novamente percebe-se uma redução do valor de *R-Sq (adj)*, neste caso, de 3,30% em relação ao modelo (17).

Tabela 16 – Resultado da função *Best Subsets* do Minitab para o conjunto das quinze variáveis.

Vars	R-Sq	R-Sq (adj)	Mallows Cp	S	R													
					I	S	A	C	N	K	A	K	B	R	I	S	C	O
1	63,5	63,40	63,20	634,70	2132,70													
2	81,7	81,60	81,40	138,30	1512,30				X									X
3	83,3	83,10	82,90	97,30	1448,00				X						X			X
4	84,3	84,10	83,90	71,20	1404,80	X			X						X		X	X
5	84,8	84,60	84,30	59,60	1384,30	X			X						X	X	X	
6	85,3	85,00	84,60	48,70	1364,60	X			X	X			X	X				X
7	85,6	85,40	85,00	40,70	1349,50	X	X		X			X			X	X	X	
8	86,0	85,70	85,30	33,10	1334,70	X	X		X	X		X			X	X	X	
9	86,4	86,10	85,60	23,40	1316,10	X	X		X	X	X	X	X					X
10	86,6	86,30	85,80	18,90	1306,40	X	X	X	X	X	X	X	X					X
11	86,9	86,50	86,00	14,50	1296,70	X	X		X	X	X	X	X		X	X	X	X
12	87,0	86,60	86,00	13,30	1292,70	X	X		X	X	X	X	X		X	X	X	X
13	87,1	86,60	86,00	13,50	1291,30	X	X	X	X	X	X	X	X		X	X	X	X
14	87,1	86,60	85,90	14,50	1291,20	X	X	X	X	X	X	X	X		X	X	X	X
15	87,1	86,60	85,90	16,00	1292,20	X	X	X	X	X	X	X	X		X	X	X	X

Na Tabela 16 é possível verificar os coeficientes do modelo (12) de acordo com a simulação executada pelo Minitab.

Tabela 17 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo com 12 variáveis.

Term	Coef	SE Coef	T-Value	P-Value	VIF
DJI	1,435	0,41	3,47	0,001	26,94
HSENG	1,445	0,19	7,67	0,000	4,79
IPCMEX	0,254	0,07	3,42	0,001	3,17
NASDAQC	-7,630	1,49	-5,14	0,000	78,44
SHNGYC	8,430	1,78	4,74	0,000	8,85
EURO	-8703,000	2322,00	-3,75	0,000	59,77
SELIC	1287,000	264,00	4,87	0,000	30,24
LIBRA	3671,000	2054,00	1,79	0,075	91,70
IENE	541841,000	209945,00	2,58	0,010	10,67
TRIGO	-8,160	3,27	-2,50	0,013	6,62
AÇUCAR NY	760,000	145,00	5,25	0,000	3,99
EMBI+ Risc	-35,590	7,46	-4,77	0,000	10,20
Constant	-13350,000	8266,00	-1,62	0,107	0,00

Nesse modelo foi possível perceber uma diminuição do *VIF* mais acentuada que no modelo anterior, apenas a variável TRIGO apresentou alta desse parâmetro, e mesmo assim uma alta não significativa de 0,26. Com esse comportamento, é esperado uma melhora do modelo (12) em relação ao modelo (17).

Ainda na Tabela 17, percebe-se que todas as variáveis são significativas para o modelo (12). De acordo com os valores de *T-Value* e *P-Value*, apenas a constante apresenta um nível de significância baixo, dado o seu *P-Value*, mas mesmo assim será mantida no modelo.

Com base na Tabela 17, foi elaborada equação matemática a seguir, que representa o modelo (12).

$$\begin{aligned}
 IBOVESPA_t = & -13350 + 1,435 DJI_t + 1,445 HSENG_t + 0,254 IPCMEX_t - 7,63 NASDAQC_t + \\
 & 8,43 SHNGYC_t - 8703 EURO_t + 1287 SELIC_t + 3671 LIBRA_t + 541841 IENE_t - 8,16 TRIGO_t + \\
 & 760 AÇÚCAR NY_t - 35,59 EMBI RISCO BRASIL_t \quad (4-5)
 \end{aligned}$$

A Figura 47 apresenta o gráfico do modelo (12) onde é possível verificar a

aderência do modelo aos dados observados, bem como sua capacidade de previsão. Novamente no período utilizado para validação, percebe-se um descolamento da curva vermelha em relação à curva azul.

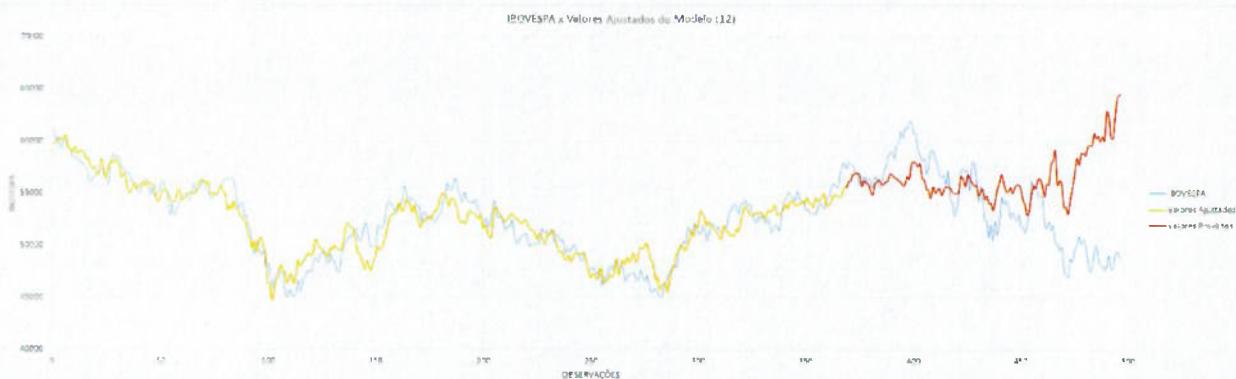


Figura 47 – Gráfico para comparação entre os dados observados, dados ajustados e previstos pelo modelo (12).

4.2.3.1 Análise dos resíduos e modelo ARMA (p, q)

Da mesma forma que para os modelos anteriores, são apresentados os gráficos dos resíduos nas Figura 48 e 49.

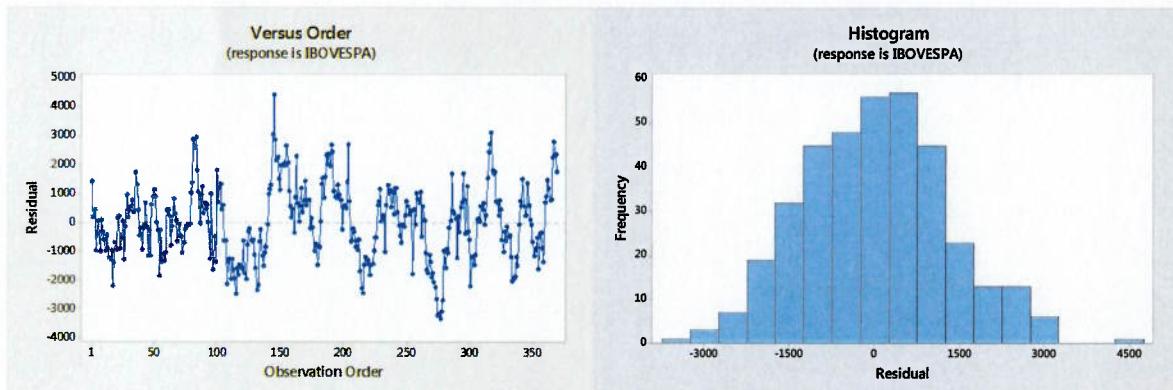


Figura 48 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (12).

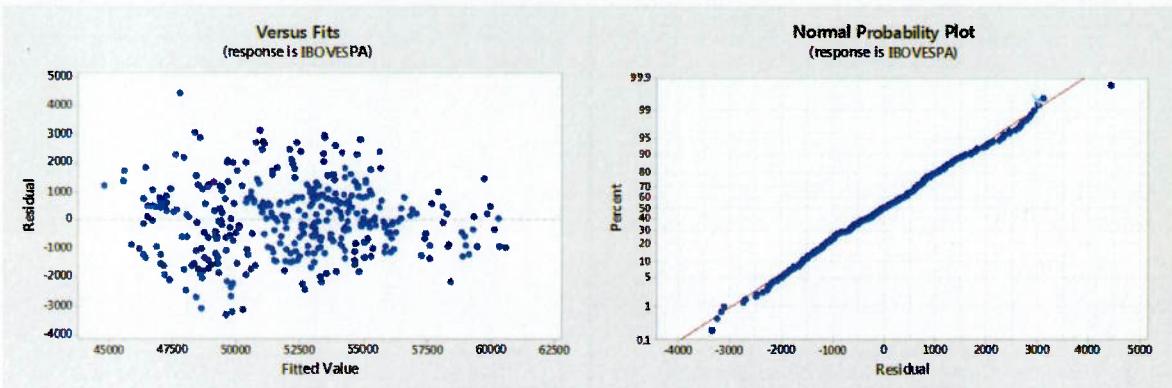


Figura 49 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.

Na Figura 50 são apresentados os gráficos que auxiliam na determinação do modelo ARMA, os gráficos de autocorrelação e autocorrelação parcial.

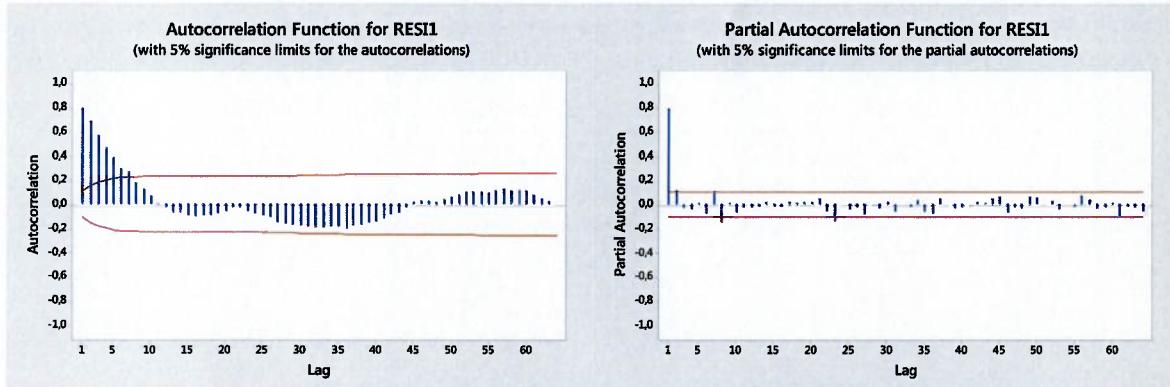


Figura 50 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (12).

Foi elaborado o modelo AR de ordem *um* para o conjunto de resíduos do modelo (12), representado pela Tabela 18. E da mesma forma que ocorreu para os modelos AR anteriores, a constante não é significativa e pode ser desconsiderada.

Tabela 18 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (12).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,8089	0,0310	26,13	0,0000
Constant	6,8600	39,2600	0,17	0,8610

O modelo AR(1) para os resíduos do modelo (12), pode ser representado matematicamente conforme a equação a seguir.

$$Res_t = 0,8089 Res_{t-1} \quad (4-6)$$

Uma vez que foi elaborado o modelo AR(1) (4-6), o mesmo foi adicionado ao modelo (12) e consequentemente foi possível perceber uma melhora na aderência das previsões do modelo, representada pelo gráfico da Figura 51.

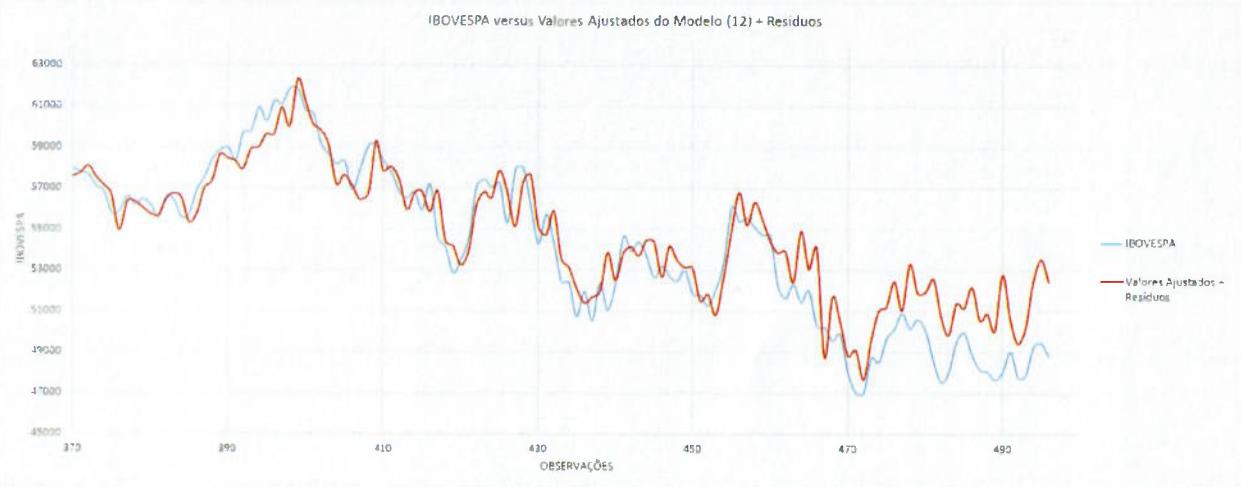


Figura 51 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (12) com a adição do modelo auto regressivo dos resíduos (4-6).

Na Figura 52 é possível perceber que o modelo (12) apresenta uma melhora de 4,87% em relação ao modelo (17). Esse resultado é explicado pela diminuição da multicolinearidade entre as variáveis explanatórias e o aumento do nível de significância das variáveis para o modelo.

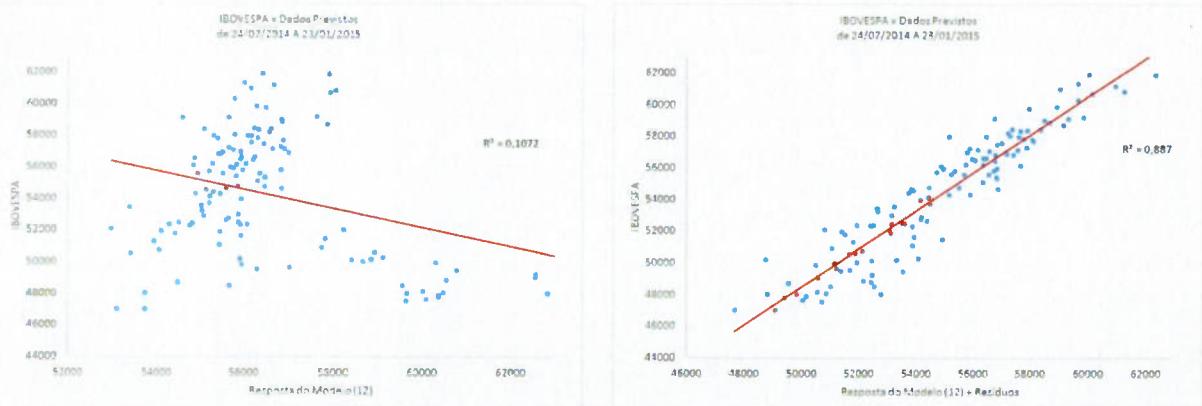


Figura 52 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (12) antes e depois da adição do modelo auto regressivo dos resíduos (4-6).

4.2.4 Simulação com o conjunto de dez variáveis

Na última simulação foram desconsideradas mais cinco variáveis, chegando-se a um conjunto de dez variáveis, mostradas na Tabela 19.

Tabela 19 – Correlações dos dez índices mais relevantes com o IBOVESPA.

ÍNDICE	FATOR DE CORRELAÇÃO	ÍNDICE	FATOR DE CORRELAÇÃO
IPCMEX	0,7202	LIBRA	-0,5898
SHNGYC	0,7124	IENE	-0,6336
OURO	0,6404	EURO	-0,6765
AÇÚCAR NY	0,6055	DÓLAR	-0,7459
HSENG	0,5846	EMBI+Risco Brasil	-0,7970

Após executar a simulação com este conjunto de variáveis, o Minitab apresentou o resultado que pode ser observado na Tabela 20. Analisando o resultado, percebe-se que o melhor modelo contém apenas nove variáveis, denominado modelo (9) daqui em diante, e novamente percebe-se uma redução do valor de $R\text{-}Sq\text{ (adj)}$, neste caso, de 1,27% em relação ao modelo (12).

Tabela 20 – Resultado da função *Best Subsets* do Minitab 17 para o conjunto das dez variáveis.

Na Tabela 21 é possível verificar os coeficientes do modelo (9) de acordo com a simulação executada pelo Minitab.

Tabela 21 – Coeficientes da regressão múltipla calculados pelo Minitab 17 para o modelo com 9 variáveis.

Term	Coef	SE Coef	T-Value	P-Value	VIF
HSENG	1,585	0,18	8,85	0,000	4,02
IPCMEX	0,268	0,07	3,62	0,000	2,93
SHNGYC	3,760	1,41	2,67	0,008	5,14
DOLAR	-8497,000	3959,00	-2,15	0,033	58,54
EURO	-10134,000	2343,00	-4,33	0,000	56,51
LIBRA	5719,000	1337,00	4,28	0,000	36,08
IENE	771366,000	193918,00	3,98	0,000	8,46
AÇUCAR NY	347,000	127,00	2,72	0,007	2,87
EMBI+ Risco-Brasil	-29,240	7,77	-3,76	0,000	10,30
Constant	9209,000	5847,00	1,58	0,116	0,00

E a partir da Tabela 19, chega-se ao modelo matemático que representa o modelo (9):

$$IBOVESPA_t = 9209 + 1,585 HSENG_t + 0,268 IPCMEX_t + 3,76 SHNGYC_t - 8497 DÓLAR_t - 10134 EURO_t + 5719 LIBRA_t + 771366 IENE_t + 347 AÇÚCAR NY_t - 29,24 EMBI RISCO BRASIL_t \quad (4-7)$$

A Figura 53 apresenta o gráfico do modelo (9) bem como sua aderência aos dados observados.



Figura 53 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (9).

4.2.4.1 Análise dos resíduos e modelo ARMA (p, q)

Nas Figuras 54 e 55 estão os gráficos dos resíduos do modelo (9). Igualmente aos outros modelos, percebe-se que a série com os resíduos do modelo (9) também é estacionária, não apresentando nenhuma tendência.

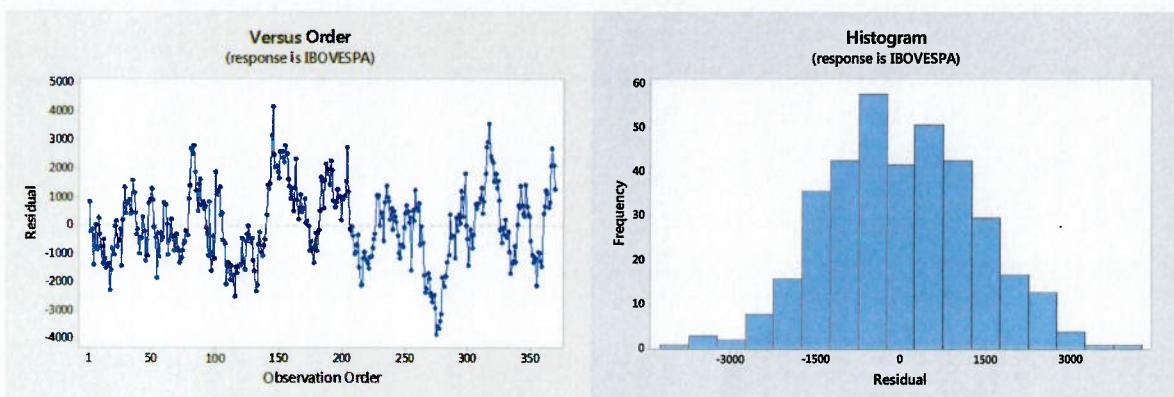


Figura 54 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (9).

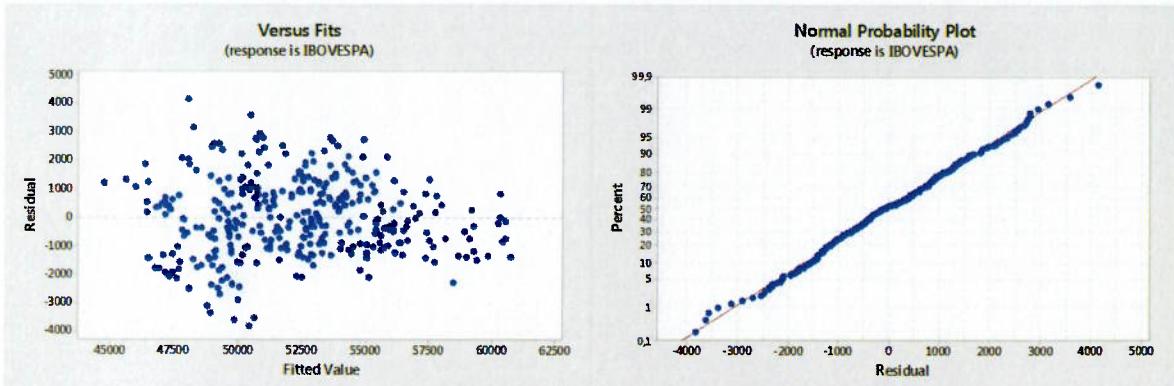


Figura 55 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos do modelo (9).

Na Figura 56 verificam-se os gráficos que auxiliam na determinação do modelo ARMA, os gráficos de autocorrelação e autocorrelação parcial.

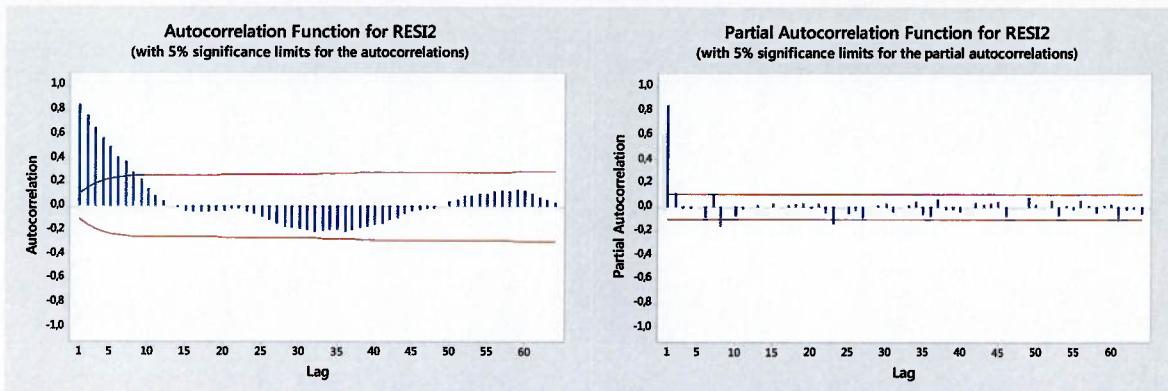


Figura 56 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (9).

Foi elaborado o modelo AR de ordem *um* para o conjunto de resíduos do modelo (9), representado pela Tabela 22. E da mesma forma que ocorreu para os modelos AR anteriores, a constante não é significativa e pode ser desconsiderada.

Tabela 22 – Modelo auto regressivo AR(1) para a série dos resíduos do modelo (12).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,8514	0,0275	30,94	0,0000
Constant	4,6800	36,3900	0,13	0,8980

Após a análise dos gráficos anteriores e a simulação do Minitab, chegou-se ao seguinte modelo AR(1):

$$Res_t = 0,8514 Res_{t-1} \quad (4-8)$$

Uma vez que foi elaborado o modelo AR(1) (4-8), o mesmo foi adicionado ao modelo (9) e consequentemente foi possível perceber uma melhora na aderência das previsões do modelo, representada pelo gráfico da Figura 57.

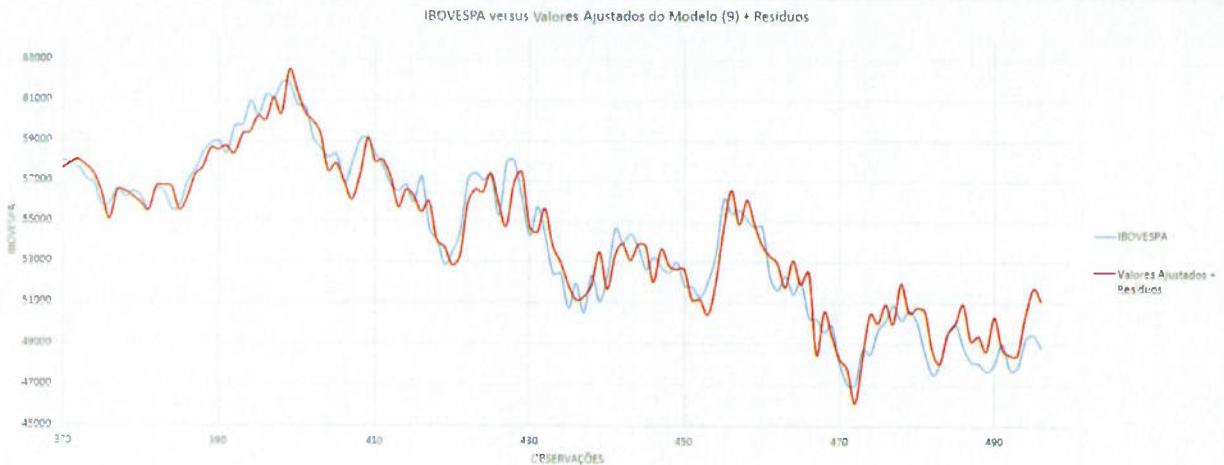


Figura 57 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (9) com a adição do modelo auto regressivo dos resíduos (4-8).

Na Figura 58 é possível perceber que o modelo (9) apresenta uma melhora do R^2 em 4,48% em relação ao modelo (12) para o período de teste.

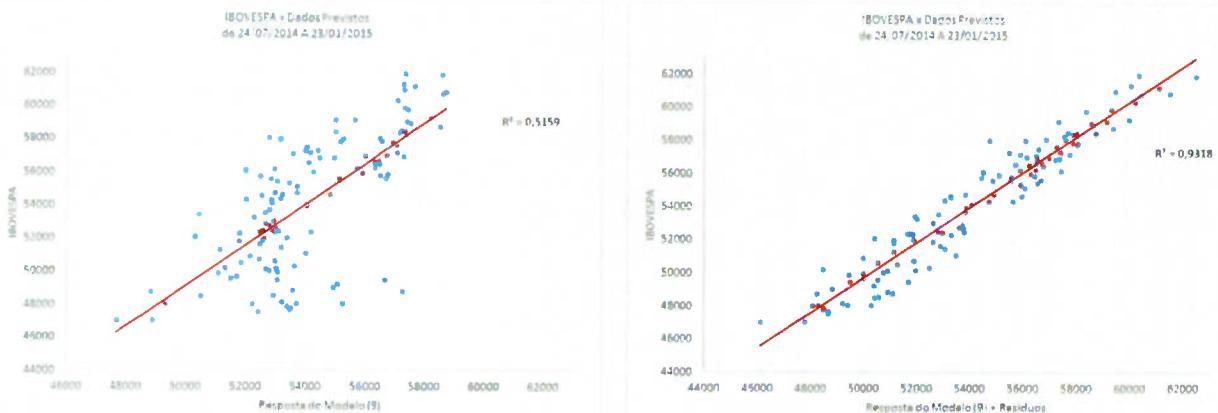


Figura 58 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (9) antes e depois da adição do modelo auto regressivo dos resíduos (4-8).

Conforme os modelos são gerados com menos variáveis, percebe-se uma variação do valor de R^2 , mas ao mesmo tempo quando os modelos são testados com dados não utilizados na modelagem é possível notar uma melhora nos resultados, conforme algumas variáveis são retiradas dos modelos. Esse fenômeno se deve à multicolinearidade e grau de significância das variáveis explanatórias. Na Tabela 23 é

possível verificar os valores do parâmetro VIF apresentados para os modelos anteriores⁵, e como ele decresce conforme os modelos vão sendo elaborados com menos variáveis. Esse comportamento fez com que os modelos com menos variáveis fizessem previsões melhores que aqueles com mais variáveis em alguns casos

Tabela 23 – Comparação do valor do parâmetro VIF (multicolinearidade) das variáveis do modelo (9), com os modelos anteriores que também contém as mesmas variáveis.

	MODELO (18)	MODELO (17)	MODELO (12)	MODELO (9)
R ²	86,8%	83,83%	88,7%	93,18%
HSENG	9,61	7,79	4,79	4,02
IPCMEX	5,04	3,94	3,17	2,93
SHNGYC	10,56	8,99	8,85	5,14
DÓLAR	-	77,22	-	58,54
EURO	-	106,06	59,77	56,51
LIBRA	61,23	110,62	91,7	36,08
IENE	27,19	-	10,67	8,46
AÇUCAR NY	-	-	3,99	2,87
EMBI+RISCO BRASIL	14,13	-	10,2	10,3

4.2.5 Simulação de um modelo de regressão com o IBOVESPA deslocado

Após verificado os resultados dos quatro modelos anteriores, uma nova simulação é realizada com o mesmo grupo de dez variáveis da Tabela 19, porém, nessa simulação, o modelo foi calculado com a série do IBOVESPA deslocada em cinco dias em relação às demais variáveis, para que fosse possível observar a influência que valores passados das variáveis explanatórias possuem sobre o IBOVESPA.

Com o grupo de dez variáveis, segue-se o mesmo procedimento adotado para os modelos (9), (12), (17) e (18), e na Tabela 24 é apresentado o resultado do *Best Subsets* do Minitab, indicando que o melhor modelo contém apenas sete variáveis, denominado modelo (7) daqui em diante, e novamente percebe-se uma redução do valor de *R-Sq (adj)*, neste caso, de 10,29% em relação ao modelo (9), uma piora acentuada em relação aos modelos anteriores.

⁵ A Tabela 23 mostra apenas as variáveis do modelo (9) e qual o valor do VIF das mesmas variáveis nos modelos anteriores.

Tabela 24 – Resultado da função *Best Subsets* do Minitab para o conjunto de dez variáveis, porém com o IBOVESPA deslocado de cinco dias.

Vars	R-Sq		R-Sq		Mallows		Y I R A N G E R N R						
	R-Sq	(adj)	(pred)	Cp	S	C	B	C	I	H R	D A T P	L	
1	58,1	58,00	57,70	298,40	2242,00								X
2	71,5	71,30	71,10	88,20	1851,50				X				X
3	73,0	72,80	72,40	66,80	1805,30			X	X				X
4	74,8	74,50	74,10	40,50	1746,90	X		X	X				X
5	75,9	75,60	75,10	24,80	1710,10			X	X	X	X		X
6	76,6	76,30	75,80	14,70	1685,30		X	X	X	X	X		X
7	77,1	76,70	76,10	9,10	1670,00	X		X	X	X	X	X	X
8	77,2	76,70	76,00	10,20	1670,20	X		X	X	X	X	X	X
9	77,3	76,70	76,00	10,10	1667,90	X	X	X	X	X	X	X	X
10	77,4	76,80	75,80	11,00	1667,60	X	X	X	X	X	X	X	X

Na Tabela 25 estão os valores dos coeficientes do modelo (7), de acordo com a simulação executada pelo Minitab.

Tabela 25 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo com 7 variáveis.

Term	Coef	SE Coef	T-Value	P-Value	VIF
HSENG	1,492	0,16	9,61	0,000	1,95
IPCMEX	0,248	0,09	2,77	0,006	2,77
EURO	-13193,000	2521,00	-5,23	0,000	42,21
LIBRA	5897,000	1468,00	4,02	0,000	28,09
IENE	946355,000	194412,00	4,87	0,000	5,48
AÇUCAR NY	541,000	128,00	4,22	0,000	1,87
EMBI+ Risco-Brasil	-53,430	8,92	-5,99	0,000	8,75
Constant	7786,000	5050,00	1,54	0,124	0,00

E a partir da Tabela 25, chega-se ao modelo matemático que representa o modelo (7).

$$IBOVESPA_t = 7786 + 1,492 HSENG_{t-5} + 0,248 IPCMEX_{t-5} - 13193 EURO_{t-5} + 5897 LIBRA_{t-5} + 946355 IENE_{t-5} + 541 AÇÚCAR NY_{t-5} - 53,43 EMBI RISCO BRASIL_{t-5} \quad (4-9)$$

Na Tabela 26 está uma comparação dos valores do parâmetro *VIF* entre os modelos (7) e (9), e é possível observar que houve uma diminuição da multicolinearidade.

Tabela 26 – Comparação do parâmetro *VIF* entre os modelos (7) e (9).

	MODELO (9)	MODELO (7)
HSENG	4,02	1,95
IPCMEX	2,93	2,77
EURO	56,51	42,21
LIBRA	36,08	28,09
IENE	8,46	5,48
AÇÚCAR NY	2,87	1,87
EMBI+RISCO BRASIL	10,3	8,75

Na Figura 59 se observa a representação gráfica do modelo (7) durante o período dos dados utilizados para a elaboração do mesmo e o período de previsão.

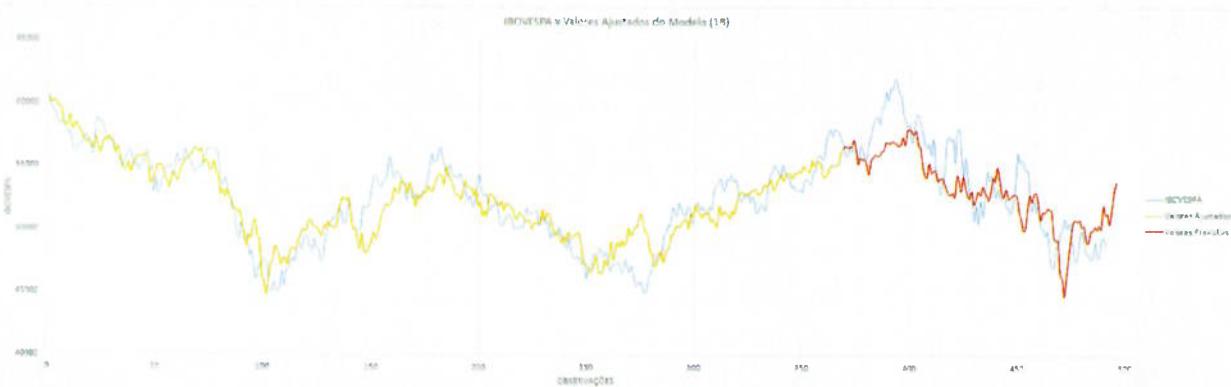


Figura 59 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (7).

4.2.5.1 Análise dos resíduos e modelo ARMA (p, q)

Da mesma forma que para os modelos anteriores, nas Figura 60 e 61 estão os gráficos da análise dos resíduos do modelo (7). No gráfico que mostra os resíduos ao longo do tempo, percebe-se alguns picos nas observações próximas de 100, 150 e 275.

No gráfico de dispersão dos resíduos versus os valores ajustados, também é possível observar que a distribuição dos dados não está uniforme na região mais à esquerda. Com essas observações, percebe-se que a série de resíduos do modelo (7) não apresenta a mesma estacionariedade dos resíduos dos modelos anteriores, mas ainda apresenta uma similaridade com uma série estacionária de distribuição normal.

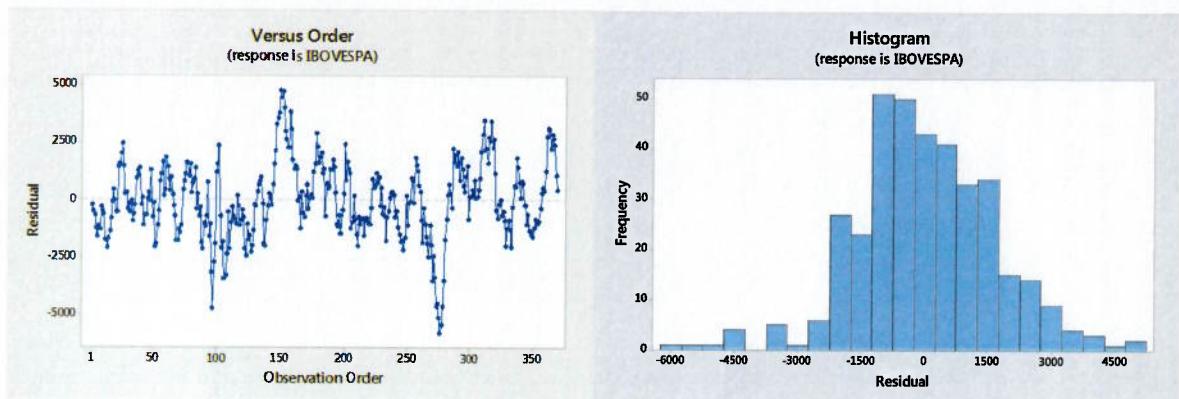


Figura 60 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (7).

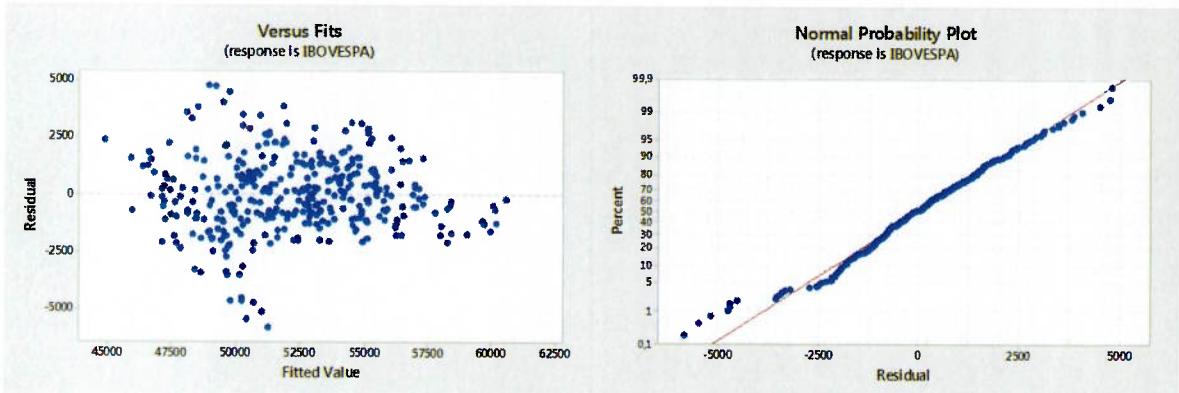


Figura 61 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos do modelo (7).

Na Figura 62 são apresentados os gráficos que auxiliam na determinação do modelo ARMA, os gráficos de autocorrelação e autocorrelação parcial. Assim como ocorreu para todos os outros casos, o modelo que melhor descreve os resíduos é um auto regressivo de ordem *um*, mas como o modelo de regressão múltipla está considerando um deslocamento de cinco dias do IBOVESPA em relação às outras variáveis, o modelo AR elaborado nesse caso tem ordem *cinco* para que se possa

verificar a possibilidade de melhorar as previsões de valores futuros do IBOVESPA. A Tabela 27 apresenta os coeficientes calculados para o modelo AR(5).

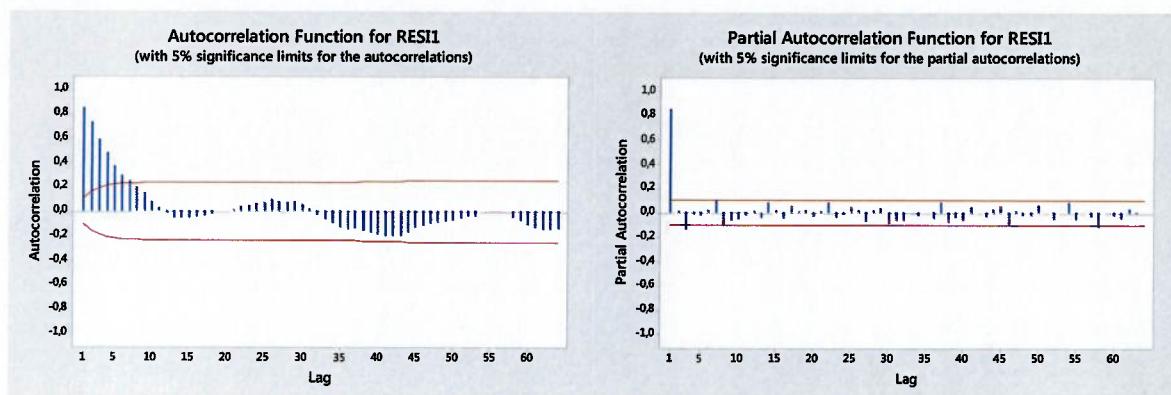


Figura 62 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (7).

Tabela 27 – Modelo auto regressivo AR(5) para a série dos resíduos do modelo (12).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,8449	0,0525	16,10	0,0000
AR 2	0,1326	0,0687	1,93	0,0540
AR 3	-0,1261	0,0689	-1,83	0,0680
AR 4	0,0061	0,0688	0,09	0,9300
AR 5	-0,0233	0,0528	-0,44	0,6590
Constant	-0,2900	44,1200	-0,01	0,9950

Assim como nos demais casos, a constante do modelo será descartada por não ser significativa para o cálculo das previsões, e é elaborada a seguinte equação matemática que representa os resíduos do modelo (7).

$$Res_t = 0,8449 Res_{t-1} + 0,1326 Res_{t-2} - 0,1261 Res_{t-3} + 0,0061 Res_{t-4} - 0,0233 Res_{t-5} \quad (4-10)$$

Como o modelo de regressão foi elaborado com o IBOVESPA deslocado em cinco dias para frente, algumas análises podem ser realizadas em relação à resposta que se deseja. Pode-se considerar o cálculo do IBOVESPA para $t, t + 1, t + 2, t + 3, t + 4$ e $t + 5$. Considera-se que para $t + 5$ os valores das variáveis explanatórias sejam conhecidos.

Considerando a resposta do IBOVESPA em t , tem-se a seguinte equação matemática e a representação gráfica para o período de testes na Figura 63

$$\begin{aligned} IBOVESPA_t = 7786 + 1,492 \cdot HSENG_{t-5} + 0,248 \cdot IPCMEX_{t-5} - 13193 \cdot EURO_{t-5} + 5897 \cdot LIBRA_{t-5} + \\ 946355 \cdot IENE_{t-5} + 541 \cdot AÇÚCAR\ NY_{t-5} - 53,43 \cdot EMBI\ RISCO\ BRASIL_{t-5} + [Res_t = 0,8449 \cdot Res_{t-1} + \\ 0,1326 \cdot Res_{t-2} - 0,1261 \cdot Res_{t-3} + 0,0061 \cdot Res_{t-4} - 0,0233 \cdot Res_{t-5}] \end{aligned} \quad (4-11)$$

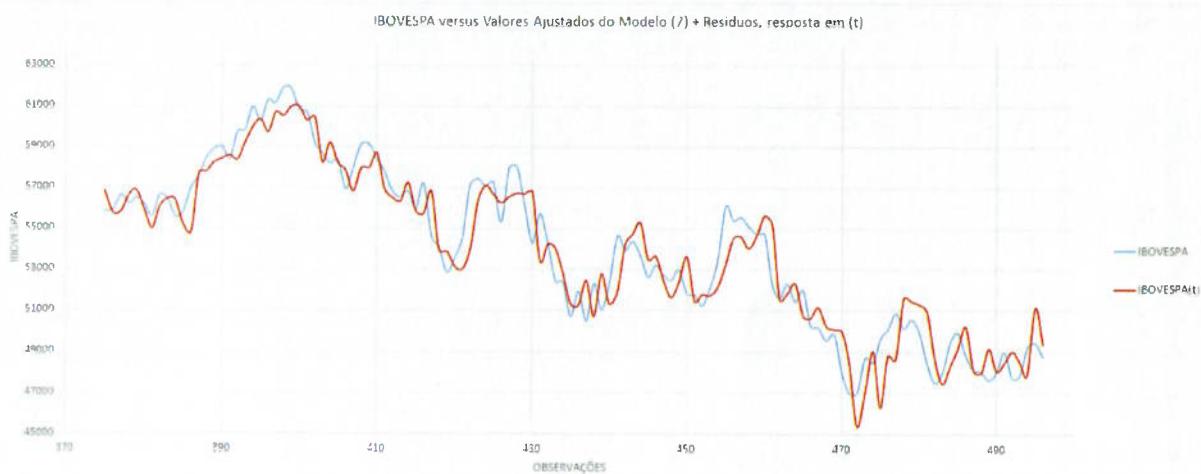


Figura 63 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-11).

Através da Figura 64 se verifica como a resposta do modelo de regressão múltipla melhora com a adição do modelo dos resíduos.

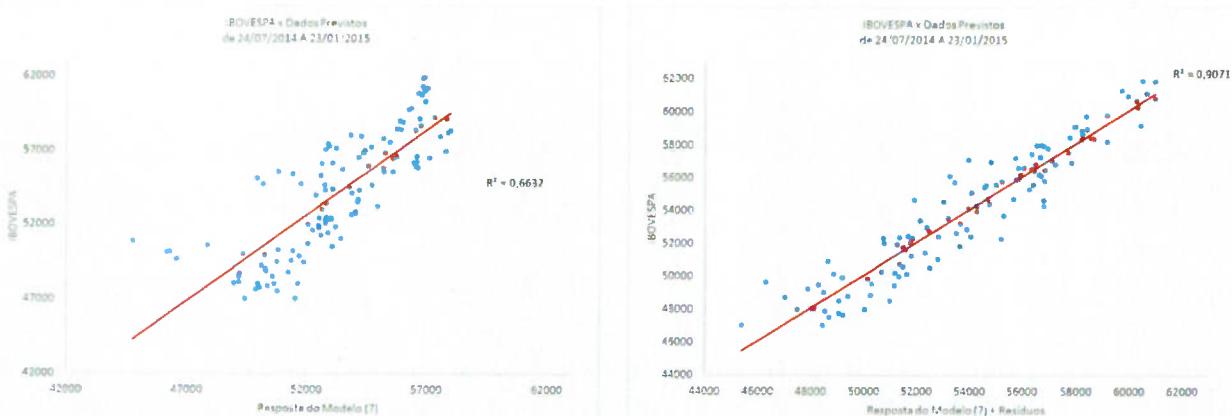


Figura 64 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-11).

Considerando a resposta do IBOVESPA em $t + 1$, tem-se a seguinte equação

matemática e a representação gráfica para o período de testes na Figura 65.

$$\begin{aligned}
 IBOVESPA_{t+1} = & 7786 + 1,492 \cdot HSENG_{t-4} + 0,248 \cdot IPCMEX_{t-4} - 13193 \cdot EURO_{t-4} + 5897 \cdot LIBRA_{t-4} + \\
 & 946355 \cdot IENE_{t-4} + 541 \cdot AÇÚCAR\ NY_{t-4} - 53,43 \cdot EMBI\ RISCO\ BRASIL_{t-4} + [Res_{t+1} = 0,1326 \cdot Res_{t-1} - \\
 & 0,1261 \cdot Res_{t-2} + 0,0061 \cdot Res_{t-3} - 0,0233 \cdot Res_{t-4}] \quad (4-12)
 \end{aligned}$$

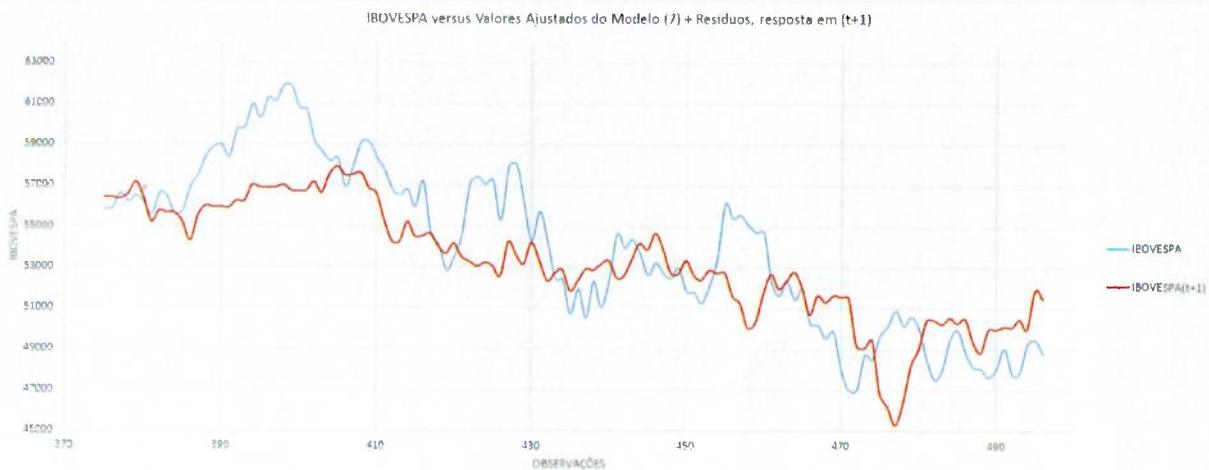


Figura 65 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-12).

A Figura 66 apresenta como a resposta do modelo de regressão múltipla piora quando se tenta prever o IBOVESPA para $t + 1$.

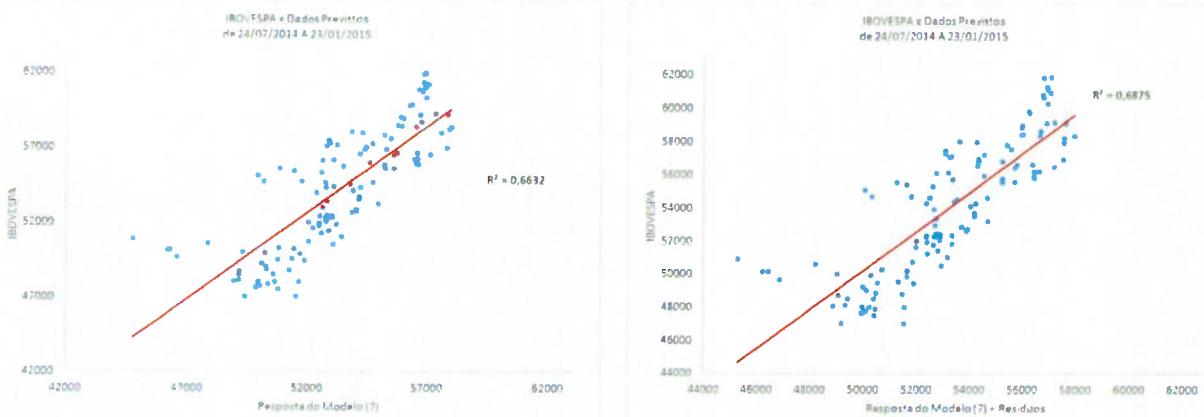


Figura 66 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-12).

Considerando a resposta do IBOVESPA em $t + 2$, tem-se a seguinte equação matemática e a representação gráfica para o período de testes na Figura 67.

$$\begin{aligned} IBOVESPA_{t+2} = & 7786 + 1,492 \cdot HSENG_{t-3} + 0,248 \cdot IPCMEX_{t-3} - 13193 \cdot EURO_{t-3} + 5897 \cdot LIBRA_{t-3} + \\ & 946355 \cdot IENE_{t-3} + 541 \cdot AÇÚCAR\ NY_{t-3} - 53,43 \cdot EMBI\ RISCO\ BRASIL_{t-3} + [Res_{t+2} = -0,1261 \cdot Res_{t-1} + \\ & 0,0061 \cdot Res_{t-2} - 0,0233 \cdot Res_{t-3}] \end{aligned} \quad (4-13)$$

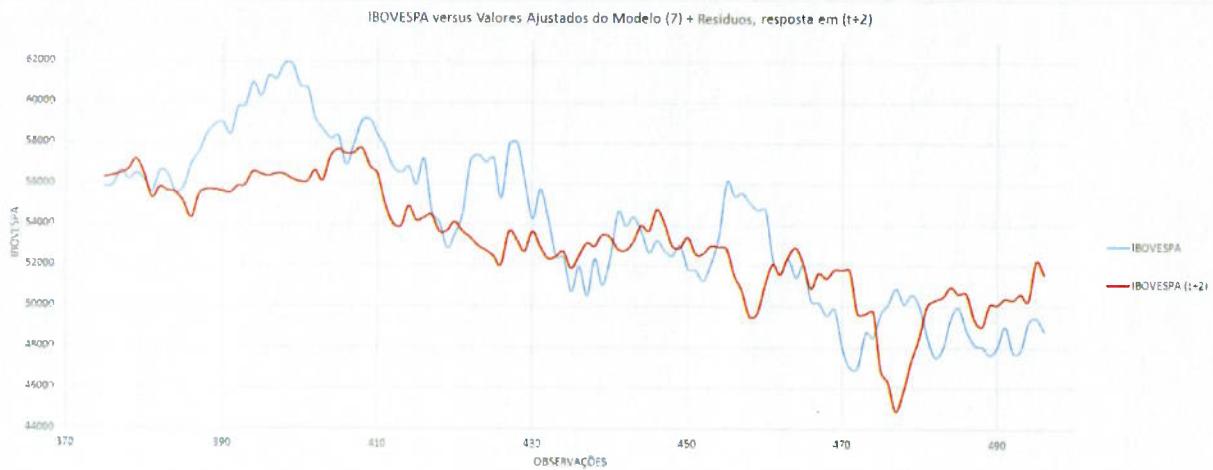


Figura 67 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-13).

A Figura 68 apresenta o comportamento da resposta do modelo de regressão múltipla quando se tenta prever o IBOVESPA para $t + 2$.

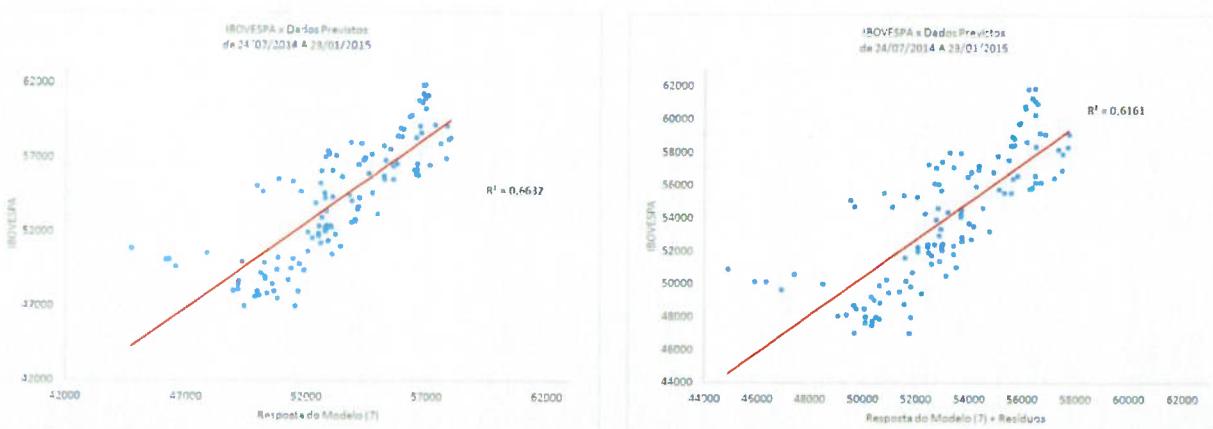


Figura 68 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-13).

Considerando a resposta do IBOVESPA em $t + 3$, tem-se a seguinte equação matemática e a representação gráfica para o período de testes na Figura 69.

$$\begin{aligned} IBOVESPA_{t+3} = & 7786 + 1,492 \cdot HSENG_{t-2} + 0,248 \cdot IPCMEX_{t-2} - 13193 \cdot EURO_{t-2} + 5897 \cdot LIBRA_{t-2} + \\ & 946355 \cdot IENE_{t-2} + 541 \cdot AÇÚCAR\ NY_{t-2} - 53,43 \cdot EMBI\ RISCO\ BRASIL_{t-2} + [Res_{t+3} = 0,0061 \cdot Res_{t-1} - \\ & 0,0233 \cdot Res_{t-2}] \end{aligned} \quad (4-14)$$



Figura 69 – Cotação histórica do IBOVESPA e dados previstos pela equação (4-14).

A Figura 70 apresenta o comportamento da resposta do modelo de regressão múltipla quando se tenta prever o IBOVESPA para $t + 3$. Percebe-se que o modelo de resíduos não afeta mais o modelo de regressão (4-9).

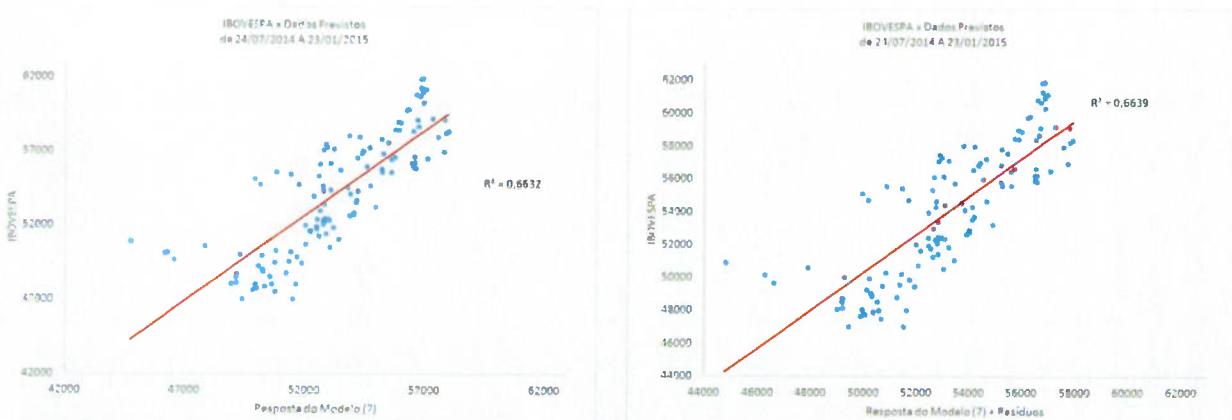


Figura 70 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pela equações (4-9) e (4-14).

As previsões do IBOVESPA para $t + 4$ e $t + 5$ apresentam o mesmo comportamento da previsão $t + 3$. Isso acontece porque o modelo auto regressivo de ordem *cinco* dos resíduos, que foi adicionado ao modelo principal, se torna impróprio a partir do momento que se decide prever o IBOVESPA para $t + 2$, e esse comportamento se deve ao fato dos termos de terceira, quarta e quinta ordem do modelo AR não serem estatisticamente significantes, conforme indicado na Tabela 27.

4.2.6 Simulação de um modelo de regressão com variáveis deslocadas

Por fim, é elaborado um modelo que possui algumas variáveis deslocadas em relação ao IBOVESPA. Esse deslocamento é feito em forma de atraso das variáveis explanatórias, de modo que a correlação entre essas variáveis e o IBOVESPA fique mais forte, positivamente ou negativamente.

Essa análise parte do conjunto de vinte e seis variáveis, comparando atrasos de t a $t - 5$ em relação ao IBOVESPA, mas para alguns casos o atraso não melhorou a correlação com o índice, como pode ser observado na Tabela 28.

Tabela 28 – Comparação das correlações de cada variável após o atraso em relação ao IBOVESPA.

ÍNDICE	DEPOIS	ANTES	ATRASO	ÍNDICE	DEPOIS	ANTES	ATRASO
SHNGYC	0,6960	0,6960	t	CAC40	-0,1990	-0,1920	$t-5$
IPCMEX	0,6896	0,6975	$t-1$	DAX30	-0,2857	-0,2540	$t-5$
AÇÚCAR NY	0,6114	0,5895	$t-5$	PETRÓLEO	-0,2869	-0,2869	t
OURO	0,6096	0,6096	t	SP500	-0,3105	-0,2853	$t-5$
HSENG	0,5685	0,5685	t	DJI	-0,3617	-0,3180	$t-5$
TRIGO	0,4242	0,3366	$t-5$	SELIC	-0,3620	-0,3620	t
PESO ARGENTINO	0,0985	0,0985	t	NASDAQC	-0,3783	-0,3628	$t-5$
SOJA	0,0976	0,0719	$t-5$	NIKKEI225	-0,4367	-0,3950	$t-5$
CAFÉ	0,0497	-0,0162	$t-5$	LIBRA	-0,5776	-0,5776	t
MERVAL	-0,0498	-0,0484	$t-5$	IENE	-0,6524	-0,6524	t
IBEX35	-0,1126	-0,1085	$t-3$	EURO	-0,6695	-0,6695	t
FTSEMIB	-0,1410	-0,1410	t	DÓLAR	-0,7357	-0,7357	t
SMI	-0,1854	-0,1403	$t-5$	EMBI+Risco-Brasil	-0,7841	-0,7822	$t-1$

Com o intuito de manter o modelo o mais simples possível e não tornar este trabalho repetitivo, parte-se do pressuposto que o conjunto das dez variáveis mais

correlacionadas podem gerar o melhor modelo, assim como ocorreu para o modelo (9) nas análises anteriores.

Portanto, desprezando-se as variáveis menos relevantes em termos de correlação com o IBOVESPA, chega-se ao mesmo conjunto de dez variáveis que é apresentado na Tabela 19. Na Tabela 29 estão listadas as variáveis que são utilizadas nessa nova simulação e suas respectivas correlações e atrasos.

Tabela 29 – Lista das dez variáveis mais relevantes em termos de correlação com o IBOVESPA.

ÍNDICE	DEPOIS	ANTES	ATRASO
SHNGYC	0,6960	0,6960	t
IPCMEX	0,6896	0,6975	t-1
AÇÚCAR NY	0,6114	0,5895	t-5
OURO	0,6096	0,6096	t
HSENG	0,5685	0,5685	t
LIBRA	-0,5776	-0,5776	t
IENE	-0,6524	-0,6524	t
EURO	-0,6695	-0,6695	t
DÓLAR	-0,7357	-0,7357	t
EMBI+Risco-Brasil	-0,7841	-0,7822	t-1

Na Tabela 24 é apresentado o resultado do Best Subsets do Minitab, indicando que o melhor modelo contém nove variáveis, denominado modelo (9-2) daqui em diante.

Tabela 30 – Resultado da função *Best Subsets* para o conjunto das dez variáveis da Tabela 24.

	R	I	S	C	O		
	A	C	U	A	C		
	S	I	C	S	B		
	H	P	A	H	L	D	R
1	61,5	61,40	61,10	520,70	2129,20		X
2	80,7	80,60	80,40	82,80	1508,50	X	X
3	82,1	81,90	81,70	53,70	1456,30	X	X
4	83,2	83,00	82,70	30,90	1413,50	X	X
5	83,5	83,30	83,00	24,20	1399,40	X	X
6	84,1	83,80	83,40	13,80	1378,10	X	X
7	84,2	83,90	83,50	12,20	1373,20	X	X
8	84,4	84,00	83,60	11,00	1368,90	X	X
9	84,5	84,10	83,60	9,60	1364,30	X	X
10	84,6	84,10	83,50	11,00	1365,10	X	X

Na Tabela 31 estão os coeficientes do modelo (9-2) calculados pelo Minitab.

Tabela 31 – Coeficientes da regressão múltipla calculados pelo Minitab para o modelo (9-2).

Term	Coef	SE Coef	T-Value	P-Value	VIF
SHNGYC	2,990	1,40	2,14	0,033	4,64
IPCMEX	0,170	0,07	2,35	0,019	2,48
AÇUCAR NY	217,000	117,00	1,85	0,065	2,34
HSENG	1,726	0,18	9,75	0,000	3,69
LIBRA	5897,000	1363,00	4,33	0,000	35,48
IENE	948655,000	195364,00	4,86	0,000	8,29
EURO	-9226,000	2391,00	-3,86	0,000	55,78
DÓLAR	-12889,000	3679,00	-3,5	0,001	47,16
EMBI+ Risco-Brasil	-30,140	7,94	-3,8	0,000	9,80
Constant	16612,000	5269,00	3,15	0,002	0,00

Através da Tabela 31, foi elaborado o modelo matemático a seguir.

$$IBOVESPA_t = 16612 + 2,990 SHNGYC_t + 0,17 IPCMEX_{t-1} + 2017 AÇÚCAR NY_{t-5} + 1,726 HSENG_t + 5897 LIBRA_t + 948655 IENE_t - 9226 EURO_t - 12889 DÓLAR_t - 30,14 RISCO BRASIL_{t-1} \quad (4-15)$$

Na Figura 71 se observa a representação do modelo (9-2) durante o período dos dados utilizados para a elaboração do mesmo, e o período de previsão.

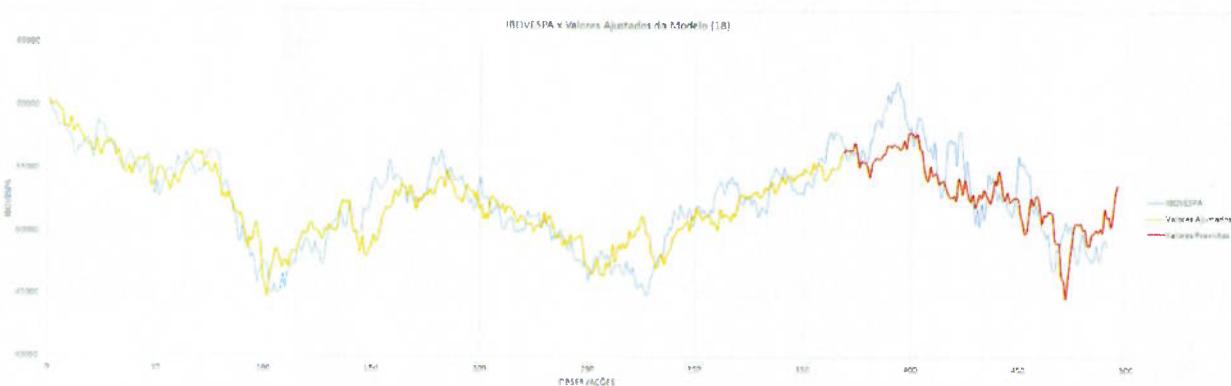


Figura 71 – Cotação histórica do IBOVESPA versus dados ajustados e previstos pelo modelo (7).

4.2.6.1 Análise dos resíduos e modelo ARMA (p, q)

Nas Figura 72 e 73 estão os gráficos da análise dos resíduos do modelo (9-2).

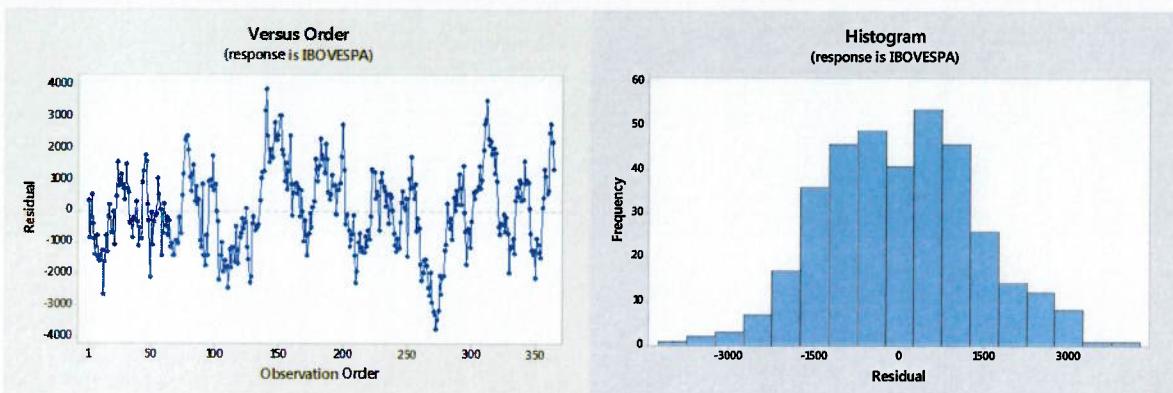


Figura 72 – Gráfico dos resíduos em relação ao tempo e histograma dos resíduos do modelo (9-2).

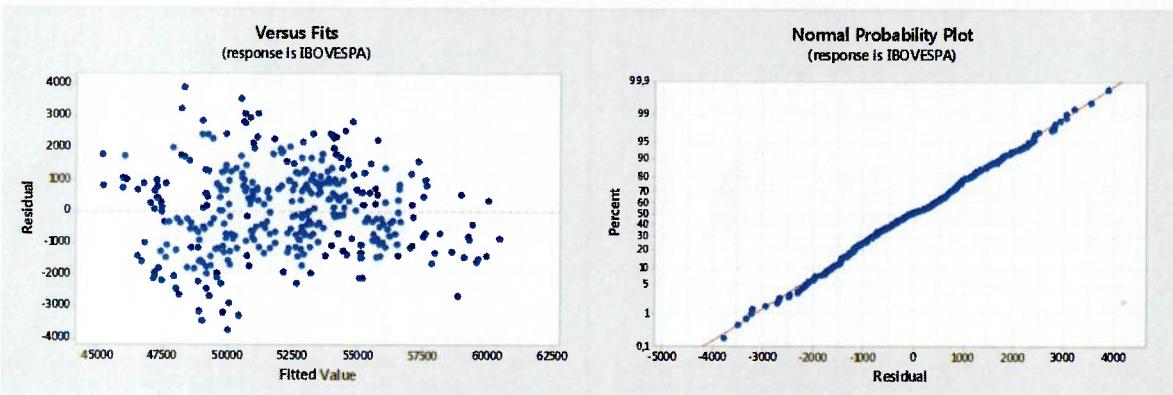


Figura 73 – Gráficos dos resíduos versus valores ajustados e normal de probabilidade dos resíduos.

Na Figura 62 são apresentados os gráficos que auxiliam na determinação do modelo ARMA, os gráficos de autocorrelação e autocorrelação parcial.

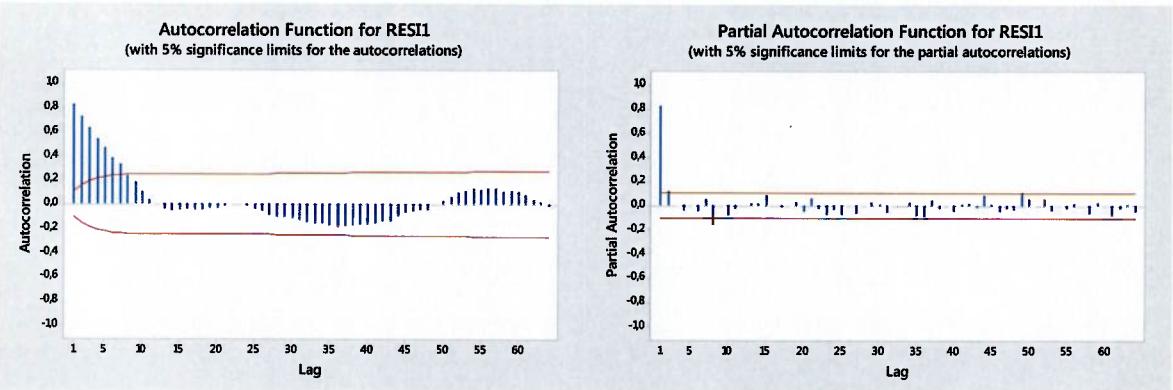


Figura 74 – Gráficos da função de autocorrelação e autocorrelação parcial dos resíduos do modelo (9-2).

Foi elaborado o modelo AR de ordem *dois* para o conjunto de resíduos do modelo (9-2), representado pela Tabela 32. E da mesma forma que ocorreu para os modelos AR anteriores, a constante não é significativa e pode ser desconsiderada.

Tabela 32 – Modelo auto regressivo AR(2) para a série dos resíduos do modelo (9-2).

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0,7289	0,0522	13,95	0,000
AR 2	0,1276	0,0524	2,43	0,015
Constant	4,2300	38,8000	0,11	0,913

O modelo AR(2) para os resíduos do modelo (9-2), pode ser representado matematicamente conforme a equação (4-16).

$$Res_t = 0,7289 Res_{t-1} + 0,1276 Res_{t-2} \quad (4-16)$$

Uma vez elaborado o modelo AR(2) (4-16), o mesmo foi adicionado ao modelo (9-2) e consequentemente foi possível perceber uma melhora na aderência das previsões do modelo, representada pelo gráfico da Figura 75.

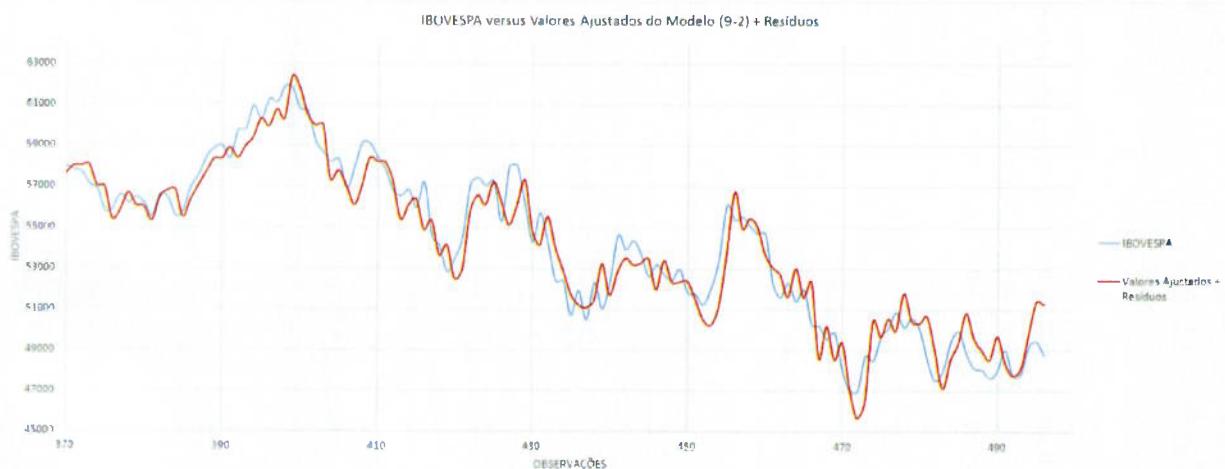


Figura 75 – Cotação histórica do IBOVESPA e dados previstos pelo modelo (9-2) mais modelo dos resíduos.

A Figura 76 apresenta o comportamento da resposta do modelo de regressão múltipla (9-2) antes e depois da adição do modelo AR dos resíduos.

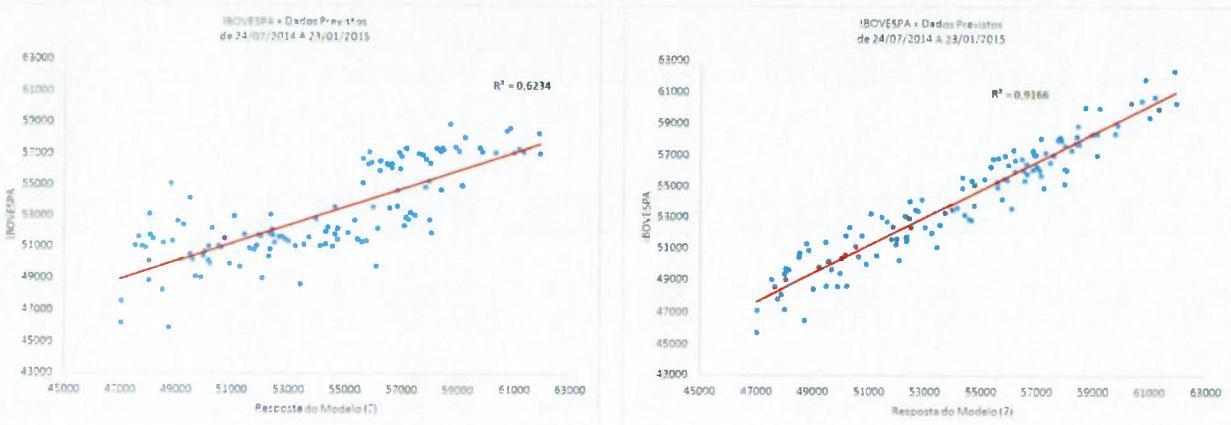


Figura 76 – Comparação dos gráficos de dispersão entre IBOVESPA e dados previstos pelo modelo (9-2) antes e depois da soma dos resíduos.

Para o modelo (9-2) não é valida a previsão do IBOVESPA para $t + 1$ porque, nesse caso, só seria possível contar com os valores das variáveis IPCMEX, ACÚCAR NY e RISCO BRASIL. Todas as outras variáveis não teriam valores conhecidos, impossibilitando, assim, a previsão do IBOVESPA.

Todas as seis simulações apresentam resultados satisfatórios em relação aos dados ajustados no período utilizado para a elaboração do modelo, mas para o período de previsão existe uma variação significativa entre eles, que pode ser ajustada, até certo ponto, somando um modelo que descreva o comportamento dos resíduos ao modelo principal. Nos casos estudados foram adicionados modelos Auto Regressivos.

Sempre que os resíduos são adicionados ao modelo principal, é possível perceber uma melhora nas previsões, mesmo que os modelos iniciais apresentem uma piora no parâmetro R^2 devido à multicolinearidade que existe entre as variáveis explanatórias.

Uma observação importante é que todos os modelos elaborados possuem as variáveis que foram selecionadas no início do capítulo 4, através da análise visual dos gráficos, indicando que esta é uma boa ferramenta para validar quais são as potenciais variáveis do modelo final.

As previsões dos modelos se mostram aderentes aos dados observados e melhoram quando os resíduos são considerados como parte da equação, mas se tornam ineficientes depois de certo número de observações, o que vai de encontro com o mencionado por (GUJARATI e PORTER, 2011), que diz que quando as séries estudadas

não são estacionárias, pode-se estudar o comportamento apenas por um certo período de tempo, e que cada conjunto de dados de séries temporais, portanto, é específico a cada episódio. Como consequência não é possível generalizar os modelos para outros períodos.

Outra deficiência desse tipo de modelo é que eles não podem ser usados para estimar valores futuros. No modelo (7) foi possível perceber isso quando se tentou estimar o valor do IBOVESPA para $t + 1$, $t + 2$ e $t + 3$, onde para a última tentativa o efeito do modelo AR sobre o modelo principal foi nulo.

A Tabela 33 apresenta uma comparação entre os seis modelos, onde é possível verificar que os dois melhores são os modelos (9) e (9-2), mas entre os dois, o que apresenta melhor aderência é o modelo (9) para respostas em t .

Tabela 33 - Comparação de R^2 entre os modelos.

	Aderência Dados Previstos
Modelo (18)	86,80%
Modelo (17)	83,83%
Modelo (12)	88,70%
Modelo (9)	93,18%
Modelo (9-2)	90,71%
Modelo (7)	90,23%

5 CONCLUSÃO

Como foi citado no início desse trabalho, um dos principais objetivos é identificar quais são as variáveis que apresentam uma relação significativa com o IBOVESPA, e que de certa forma, podem descrevê-lo no período estudado de 23/01/2013 a 23/01/2015. Diante de todos os modelos estimados, chega-se à conclusão que os índices HSENG, IPCMEX, SHNGYC, Risco Brasil e as moedas Libra, Euro e Iene têm maior representatividade sob o mercado brasileiro de ações. As variáveis que correspondem a esses índices e moedas aparecem em 80% dos modelos, sendo que o HSENG, o IPCMEX e a Libra aparecem em 100% dos casos.

Segue que esse resultado foi constatado apenas com estudos empíricos das séries temporais, não levando em consideração qualquer evento social, político ou econômico no âmbito nacional ou internacional, mas de fato, qualquer evento dessa natureza poderia influenciar no resultado de cada modelo.

Diante dos resultados obtidos, uma das atribuições que esses modelos podem ter é o de auxiliar na tomada de decisões para investimentos que possuem o IBOVESPA como referência, mas apenas em operações *Day Trade*. Como foi constatado, os modelos não são adequados para estimar valores futuros do índice IBOVESPA, um vez que as séries de resíduos não possuem memória longa e não permitem a elaboração de modelos mais consistentes.

O trabalho também leva a concluir que para fazer uso desse tipo de estudo para melhorar o desempenho de portfólios com uma perspectiva futura, seria necessário conhecer valores futuros das variáveis explanatórias, ou ter uma relação mais consistente entre valores passados dessas variáveis e o IBOVESPA.

Tão importante quanto os resultados atingidos, foram os conhecimentos colocados em prática e os adquiridos durante o período de pesquisa, como o uso da ferramenta Minitab 17 para análises de séries temporais e estatísticas.

Para trabalhos futuros, que tenham o IBOVESPA como objeto de estudo, sugere-se a elaboração de um modelo que utilize a combinação de métodos qualitativos e quantitativos através de um ambiente de inteligência artificial. Deste modo, espera-se reduzir o erro de previsão do IBOVESPA.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- FROST, J. What Are the Effects of Multicollinearity and When Can I Ignore Them? **The Minitab Blog**, 2014. Disponível em: <<http://blog.minitab.com/blog/adventures-in-statistics/what-are-the-effects-of-multicollinearity-and-when-can-i-ignore-them>>. Acesso em: 7 Novembro 2015.
- GAIO, L. E. et al. Interdependência entre os mercados mundiais de ações: uma análise de volatilidades. **Revista de Administração e Contabilidade da Unisinos**, v. 11, p. 259-274, Julho/Setembro 2014.
- GROPPÔ, D. S. **Causalidade das variáveis macroeconômicas sobre o Ibovespa**. Universidade de São Paulo. Piracicaba, p. 107. 2004.
- GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. Tradução de Denise Durante; Mônica Rosemberg e Maria Lúcia G. L. Rosa. 5^a. ed. Porto Alegre: AMGH, 2011. 924 p.
- INVESTPEDIA. O que é Free Float? **Investpedia**. Disponível em: <<http://www.investpedia.com.br/artigo/O+que+e+Free+Float.aspx>>. Acesso em: 02 Novembro 2015.
- KIYOSAKI, R. T.; LECHTER, S. L. **Pai rico, pai pobre - o que os ricos ensinam a seus filhos sobre dinheiro**. 56^a. ed. Rio de Janeiro: Campus, 2000.
- MEDEIROS, A. L. **Régressão múltipla e o modelo ARIMA na previsão da arroba do boi gordo**. UNIFEI. Itajubá, p. 110. 2006.
- MORETTIN, P. A. **Econometria financeira - Um curso em séries temporais financeiras**. 2^a. ed. São Paulo: Blucher, 2011.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2^a. ed. São Paulo: Blucher, 2006.
- NETO, S. M. D. C. **Proposição de um modelo de regressão para a previsão do índice Ibovespa**. Escola Politécnica da Universidade de São Paulo. São Paulo, p. 126. 2006.
- NUNES, M. S.; JR., N. C. A. D. C.; SEABRA, F. Co-integração e causalidade entre variáveis macroeconômicas, "risco Brasil" e retornos no mercado de ações brasileiro. **Revista de Economia e Administração**, Porto Alegre, v. 2, n. 3, p. 26-42, Julho/Setembro 2003.

PIMENTA, T. J.; HIGUCHI, R. H. Variáveis macroeconômicas e o Ibovespa - Um estudo da relação de causalidade. **REAd**, v. 14, n. 2, Maio/Agosto 2008.

SCHOR, A.; BONOMO, M. A.; PEREIRA, P. L. V. Arbitrage Pricing Theory (APT) e variáveis macroeconômicas: um estudo empírico sobre o mercado acionário brasileiro.

Revista de Economia e Administração, Rio de Janeiro, v. 1, n. 1, p. 38-63, Janeiro/Março 2002.

SILVA, F. M. D.; CORONEL, D. A. Análise da causalidade e cointegração entre variáveis macroeconômicas e o Ibovespa. **R. Adm. FACES Journal**, Belo Horizonte, v. 2, n. 3, p. 31-52, Julho/Setembro 2012.

UOL. Cotações. **UOL Economia**. Disponível em: <<http://economia.uol.com.br/cotacoes/bolsas/>>. Acesso em: 23 jan. 2015.